CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 113, No. 3 March 2003

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 113, NO. 3, MARCH 2003

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 113, NO. 3, MARCH 2003

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 113, NO. 3, MARCH 2003

CONTENTS—*Continued from preceding page*

SPEECH PRODUCTION [70]

SPEECH PERCEPTION [71]

MUSIC AND MUSICAL INSTRUMENTS [75]

BIOACOUSTICS [80]

CUMULATIVE AUTHOR INDEX

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.*

William W. Lang receives the INCE Distinguished Engineer Award

William W. Lang was awarded the Distinguished Noise Control Engineer Award of the Institute of Noise Control Engineering USA (INCE-USA) at a joint meeting of the Greater Boston Chapter of the Acoustical Society of America (ASA) and INCE-USA Cambridge, MA on 28 October (see Fig. 1).

The INCE Award is awarded no more frequently than once per year, ordinarily at an INCE national conference or international congress. The award includes a plaque on which the awardee's citation is described along with an engraved silver Paul Revere Bowl. Previous recipients have been Leo Beranek and George Maling.

William Lang is a Fellow of the Acoustical Society of America and received the ASA Silver Medal in Noise in 1984. He has served the Society as a Member of the Executive Council (1984–87) and as Treasurer (1994–98).

FIG. 1. The three recipients of the INCE-USA Distinguished Engineer Award, Leo L. Beranek (l), William W. Lang (c), and George C. Maling (r).

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.

2003

13–15 March	American Auditory Society Annual Meeting, Scottsdale, AZ [American Auditory Society, 352 Sundial Ridge Cir., Dammeron Valley, UT 84783; Tel.: 435-574-0062; Fax: 435-574-0063; E-mail: amaudsoc@aol.com; WWW: www.amauditorysoc.org].
28 April–2 May	145th Meeting of the Acoustical Society of America, Nashville, TN [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].
5–8 May	SAE Noise & Vibration Conference & Exhibition, Traverse City, MI [P. Kreh, SAE International, 755 W. Big Beaver Rd., Suite 1600, Troy, MI 48084; Fax: 724-776-1830; WWW: http://www.sae.org].
12–16 May	Symposium on Environmental Consequences of Underwater Sound (ECOUS), San Antonio, TX [www.lsro.org/ECOUS].
23–25 June	NOISE-CON 2003, Cleveland, OH [INCE Business Office, Iowa State Univ., 212 Marston Hall, Ames, IA 50011-2153; Fax: 515-294-3528; E-mail: ibo@ince.org].
27–30 July	1st Conference on Acoustic Communication by Animals, University of Maryland, College Park, MD [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/communication.html].
5–8 Oct.	IEEE International Ultrasonics Symposium, Honolulu, HI [W. D. O'Brien, Jr., Bioacoustics Research Lab., Univ. of Illinois, Urbana, IL 61801-2991; Fax: 217-244-0105; WWW: www.ieee-uffc.org].
10–14 Nov.	146th Meeting of the Acoustical Society of America, Austin, TX [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

2004

24–28 May	75th Anniversary Meeting (147th Meeting) of the Acoustical Society of America, New York, NY [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].
3–7 Aug.	8th International Conference of Music Perception and Cognition, Evanston, IL [School of Music, Northwestern Univ., Evanston, IL 60201; WWW: www.icmpc.org/conferences.html].
15–19 Nov.	148th Meeting of the Acoustical Society of America, San Diego, CA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

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

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

Some indexes are out of print as noted below.

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

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

 0001-4966/2003/113(3)/1181/2/$19.00

Volumes 21–30, 1949–1958: JASA, Contemporary Literature, and Patents. Classified by subject and indexed by author and inventor. Pp. 952. Price: ASA members $20; Nonmembers $75.
Volumes 31–35, 1959–1963: JASA, Contemporary Literature, and Patents. Classified by subject and indexed by author and inventor. Pp. 1140. Price: ASA members $20; Nonmembers $90.
Volumes 36–44, 1964–1968: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 485. Out of Print.
Volumes 36–44, 1964–1968: Contemporary Literature. Classified by subject and indexed by author. Pp. 1060. Out of Print.
Volumes 45–54, 1969–1973: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 540. Price: $20 (paperbound); ASA members $25 (clothbound); Nonmembers $60 (clothbound).
Volumes 55–64, 1974–1978: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 816. Price: $20 (paperbound); ASA members $25 (clothbound); Nonmembers $60 (clothbound).
Volumes 65–74, 1979–1983: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 624. Price: ASA members $25 (paperbound); Nonmembers $75 (clothbound).
Volumes 75–84, 1984–1988: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 625. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound).
Volumes 85–94, 1989–1993: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 736. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound).
Volumes 95–104, 1994–1998: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 632. Price: ASA members $40 (paperbound); Nonmembers $90 (clothbound).

BOOK REVIEWS

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

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

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

Auditory Sound Transmission: An Autobiographical Perspective

Jozef J. Zwislocki

Lawrence Erlbaum Associates, Mahwah, New Jersey, 2002. 419 pp. Price: $99.95 (hardcover) ISBN: 0805806792.

As suggested by the title, Dr. Zwislocki has written a book that details his view of the acoustics and mechanics of sound processing by the peripheral auditory system. The book has some autobiographical flavor to it, especially the preface where Dr. Zwislocki describes his early technical education in Switzerland, his joining the Psychoacoustics Laboratory at Harvard (at the invitation of S. S. Stevens), his move to Syracuse, and the founding of the Institute for Sensory Research. This autographical flavor contributes to, but does not set, the fundamental tone of the rest of the book. Instead, the major content of the book is organized along an anatomical and physiological gradient starting with the external ear, moving through discussions of the middle ear and cochlear mechanics, and ending in a discussion of psychoacoustics. A fundamental theme throughout the book is the use of quantitative descriptions and analytic models to describe the essence of the transformation of sound to sensation. The chapters on each of the sections of the auditory system are not written to describe the evolution in our understanding of the ear, but instead provide a view of Dr. Zwislocki's current thinking. These chapters make references to the work of many contributors to each field but concentrate on the work of the author and his immediate co-workers, with a major emphasis on cochlear mechanics. There is also a distinct effort to communicate some of the author's more recent work that has not been well described in other publications.

The first three chapters (88 pages in length, entitled ''The Path of Sound,'' ''The Outer Ear,'' and ''The Middle Ear'') review the basic structures of the ear and some of the fundamental data that describe the acoustics and mechanics of the two most peripheral sections of the auditory system. Highlights of these sections include reviews of (1) the early acoustic measurements used to describe the contribution of the substructures of the external ear to sound transmission, (2) the development of the ''Zwislocki coupler'' as a standard fixture for the calibration of audiologic earphones, (3) middle-ear impedance measurements in normal and pathologic human ears using the ''Zwislocki bridge,'' and (4) the development and testing of the structure-based analytical model of middle-ear function that has been the starting point of more elaborate modern models. The development of this model is instructive of Dr. Zwislocki's approach to the study of the auditory system. The model contains a small number of anatomically defined structures, the contribution of each being amenable to measurement. Measurements are then used to define and test the structural values used in the model. These models sometimes ignore details in structure, e.g., the three-dimensional structure of the tympanic membrane, that may be considered secondary in order to define a simple, testable analytic decription, with an endpoint being a ''sufficiently accurate representation...useful in the study of sound transmission.''

Chapters 4–6 are the heart of the book. Chapter 4 (86 pages in length and entitled ''The Cochlea Simplified by Death'') is a review of the ''long-wave'' cochlear model that was Dr. Zwislocki's first contribution to the study of the ear. The chapter starts with a description of the structure of the human auditory inner ear including the variations in structure and mechanical properties along the length of the cochlear partition. This is followed by a significant analytic section, which contains 91 equations that cover topics including (1) the presence of tonotopically distributed resonators within the cochlear partition, (2) the long-wave assumption and the resulting transmission-line equations, (3) the effect of the mass and resistance of the cochlear partition, (4) the cochlear input impedance, (5) the effect of the depth of the cochlear fluid canals and the failure of the long-wave model near the place of resonance, and (6) Békésy's paradoxical waves that traveled from the cochlear base to apex regardless of the location of the vibratory source.

Chapter 5 (105 pages in length and entitled ''Live Cochlea: Physical Constants and Fundamental Characteristics'') is a unique review (and sometimes exposition) of Dr. Zwislocki's more recent work on cochlear micromechanics. The chapter includes extensive descriptions of the methodologies used to determine the physical properties of the tectorial membrane, and record from outer-hair cells and the supporting Hensen's cells in the middle turn of the cochlea in living gerbils. A major contention of this chapter (and the rest of the book) is that the Hensen's cells are electrically coupled to the outer hair cells in a manner that allows their use as monitors of the sound-evoked potentials in outer hair cells. This contention is important because it is easier to obtain stable recordings from Hensen's cells. Such recordings will be comparators for most of the subsequent descriptions and analyses of cochlear mechanics. These recent and some newly described physiological measurements are used to assess the effective stiffness of the stereocilia of outer cells, and wave-velocity and cochlear wavelength. The last sections include a good review of modern measurements of basilar-membrane displacement in different animals and a discussion of the contribution of the frequency dependence of shear motion between the tectorial membrane and the base of the stereocilia to the frequency dependence of the outer-hair cells.

Chapter 6 (110 pages in length and entitled ''Live Cochlea: Analysis'') reviews Dr. Zwislocki's previous micromechanical model of stereocilia tectorial-membrane resonance and builds on that model to present a complete model of the auditory periphery. The chapter includes numerous tests of the model parameters and comparisons to data. The chapter concentrates on the additional filtering of cochlear mechanical events produced by the frequency dependence of the shear motion within the cochlear partition. The central point of the chapter is that an anatomically correct model exists ''whose every element corresponds to an anatomical mechanical element within the cochlea...'' and which accounts ''for most prominent characteristics of cochlear mechanics.'' The effects of varied cochlear resistance, the contribution and mechanisms for active cochlear feedback, and cochlear compression are also discussed. A unique characteristic of this chapter is a hypothesis that associates cochlear compression with an adaptation mechanism.

The last chapter (21 pages in length and entitled ''Pitch and Loudness Codes'') follows through on Dr. Zwislocki's contention (voiced in the Preface) that a sufficient analytic description of sound transmission through the auditory periphery can be the basis for research on higher levels of auditory coding including perception. The chapter discusses the association of pitch to the ''cut-off'' frequency of cochlear micromechanics (as opposed to the peak frequency), and demonstrates a correlation between loudness estimates and cochlear mechanisms.

This book is a description of an end point in a personal scientific journey, and represents a consistent view of the entire peripheral auditory system. The book is not a comprehensive review of the last 55 years of auditory science, but does note numerous intersection points between Dr. Zwislocki's work and the work of others. The book stresses quantitative descriptions and will not be easy reading for the mathematically naïve. The emphasis on quantitation and the focused mechanical and acoustic view of the auditory system presented, lead me to suggest that the book will not be useful as a primary text to new students in the field. However, it is clearly appropriate and recommended for all more senior students with a primary interest in auditory mechanics, including those senior students who have studied this field for more than 10 years.

JOHN J. ROSOWSKI
Eaton-Peabody Laboratory
Massachusetts Eye and Ear Infirmary and
Department of Otology and Laryngology
Harvard Medical School
Boston, Massachusetts 02114

 0001-4966/2003/113(3)/1191/1/$19.00

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
DAVID PREVES, *Micro-Tech Hearing Instruments, 3500 Holly Lane No., Suite 10, Plymouth, Minnesota 55447*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, Univ. of Rochester, Rochester, New York 14627*

6,439,831

43.30.Jx METHOD AND APPARATUS FOR IMPROVING EFFICIENCY AND/OR ALTERING ACOUSTIC SIGNATURE OF SURFACE AND SUBMERGED VESSELS

Wayne Ernest Conrad, Hampton, Canada
27 August 2002 (Class 415/1); filed in Canada 20 April 2000

The power supplied to the propeller of either a surface or subsurface vessel is time modulated by various electrical and mechanical means in an attempt to forestall cavitation and thereby increasing the efficiency of the propulsion system. At the same time, this power modulation can alter the characteristic noise signature of the vessel, perhaps resulting in quieter operation.—WT

6,420,816

43.35.Pt METHOD FOR EXCITING LAMB WAVES IN A PLATE, IN PARTICULAR A CONTAINER WALL, AND AN APPARATUS FOR CARRYING OUT THE METHOD AND FOR RECEIVING THE EXCITED LAMB WAVES

Igor Getman *et al.*, assignors to Endress + Hauser GmbH + Company
16 July 2002 (Class 310/313 R); filed in Germany 22 December 1999

Lamb waves are excited in the wall of a container in order to monitor the level of a liquid. Lamb waves, which are composed of transverse waves oscillating perpendicular to the surface and of longitudinal waves oscillating tangentially, here are produced by means of interdigital transducers composed of piezoelectric elements configured like interlaced fingers. These elements are attached to the container surface and are designed to generate longitudinal vibrations that couple well to the desired Lamb waves.—EEU

6,437,517

43.35.Sx METHOD AND SYSTEM FOR EXCITING AN AZIMUTHAL ACOUSTIC AND LONGITUDINAL ACOUSTIC COMBINATION MODE

Jerry Martin Kramer, assignor to Koninklijke Philips Electronics N.V.
20 August 2002 (Class 315/246); filed 22 February 2001

This system relates to reducing vertical segregation in a high-intensity discharge (HID) lamp. The device reduces this segregation through an excitation of an azimuthal and longitudinal acoustic combination mode of the HID lamp. The system includes power sources that provide current frequency signals to the lamp, exciting an azimuthal and longitudinal acoustic combination mode of the lamp, whereby color mixing is achieved within the lamp.—DRR

6,427,531

43.35.Yb ACTIVE ACOUSTIC PHASED ARRAY ANTENNA SYSTEM

Prasan Chintawongvanich, Las Cruces, New Mexico
6 August 2002 (Class 73/170.13); filed 9 November 1999

Remote measurement of atmospheric wind can be performed using radio waves (sonar), or more recently hypersonic acoustic waves. The inventor argues that adapting established sonar techniques to acoustic waves is inefficient. A new phased array system is described which uses separate transmitting and receiving arrays. The system includes an improved backscatter signal retrieval technique which is said to greatly speed up the measurement process while at the same time providing improved measurement resolution—GLA

6,416,164

43.35.Zc ACOUSTIC EJECTION OF FLUIDS USING LARGE F-NUMBER FOCUSING ELEMENTS

Richard G. Stearns and Richard N. Ellson, assignors to Picoliter, Incorporated
9 July 2002 (Class 347/46); filed 20 July 2001

This patent relates to a device for ejecting droplets from a fluid reservoir toward designated sites on a surface. The technique is said to be useful in a number of contexts, particularly in the preparation of bimolecular ar-

 0001-4966/2003/113(3)/1193/14/$19.00

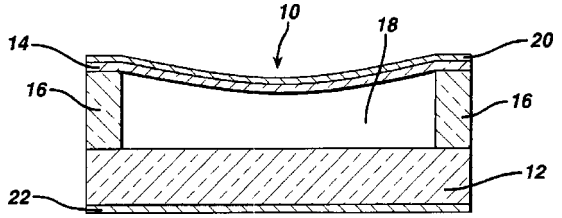
rays. Ejection of droplets from the free surface of a fluid occurs when acoustic energy of sufficient intensity is focused through the fluid onto its surface. The F-number is the ratio of the focusing-means-to-focal-point distance to the size of the aperture through which the acoustic energy passes. Arrangements with F-numbers near one are well understood, but arrangements with larger F-numbers extend the flexibility and utility of focused acoustic waves for droplet ejection. Larger F-number arrangements permit ejection of small droplets from a fluid layer of greater height and allow for greater latitude in fixing the location of the fluid surface relative to the focal plane of the acoustic beam.—EEU

6,443,901

43.38.Ar CAPACITIVE MICROMACHINED ULTRASONIC TRANSDUCERS

John Douglas Fraser, assignor to Koninklijke Philips Electronics N.V.
3 September 2002 (Class 600/459); filed 15 June 2000

An ultrasonic transducer is formed of multiple capacitive micromachined ultrasonic transducer (cMUT) cells, each comprising a charged diaphragm plate **14** capacitively adjacent to an oppositely charged base plate **12**. The diaphragm plate is distended toward the base plate by a bias charge. The base plate may include a central portion elevated toward the center of the diaphragm plate (not shown) to cause the charge of the cell to be of maximum density at the moving center of the diaphragm plate. For harmonic operation the drive pulses applied to the cells are predistorted in consideration of the inherent nonlinearity of the device to reduce harmonic contamination of the transmit signal. The cMUT cells can be fabricated by conventional semiconductor processes and hence integrated with ancillary transducer circuitry such as a bias charge regulator. The cells may also be fabricated by micro-stereolithography whereby they can be formed using a variety of polymers and other materials.—DRR



6,434,539

43.38.Hz METHOD AND APPARATUS FOR DETERMINING AND FORMING DELAYED WAVEFORMS FOR FORMING TRANSMITTING OR RECEIVING BEAMS FOR AN ACOUSTIC SYSTEM ARRAY OF TRANSMITTING OR RECEIVING ELEMENTS FOR IMAGING IN NON-HOMOGENOUS/NON-UNIFORM MEDIUMS

Harvey C. Woodsum *et al.*, assignors to Sonetech Corporation
13 August 2002 (Class 706/13); filed 20 April 1999

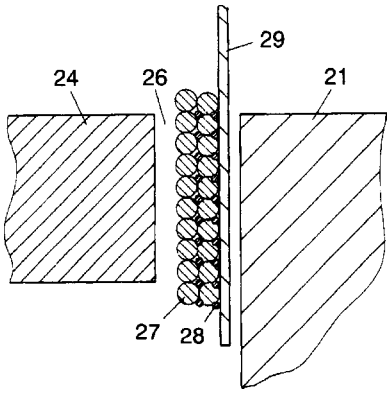
How can a beam of sound be controlled if it passes through a nonhomogenous medium such as human tissue? The inventors make use of genetic algorithms in an unusual approach to the problem. Genetic terminology (gene, chromosome, surviving population, etc.) is employed throughout the patent, including the claims. Fortunately, definitions are included where needed. Those interested in the field will find the patent worth reading.—GLA

6,421,449

43.38.Ja SPEAKER

Akinori Hasegawa *et al.*, assignors to Matsushita Electric Industrial Company, Limited
16 July 2002 (Class 381/401); filed in Japan 16 March 1999

It is well known that a loudspeaker voice coil wound with round wire rather than ribbon wastes valuable space in the magnetic gap **26**. This patent suggests that the interstices between windings **27** can be filled with smaller windings **28**. These secondary windings are then used as a detection coil for motional feedback.—GLA

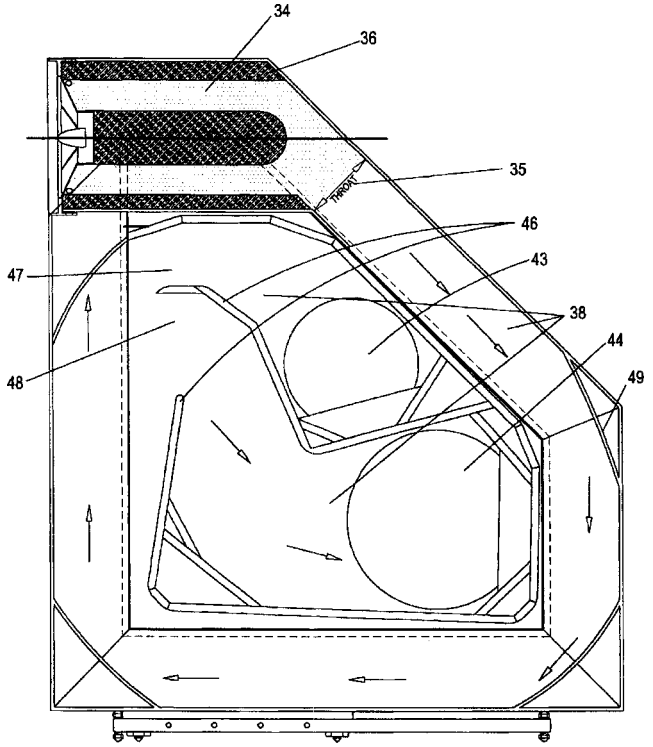


6,425,456

43.38.Ja HOLLOW SEMICIRCULARLY CURVED LOUDSPEAKER ENCLOSURE

Jacob George, assignor to Vector Transworld Corporation
30 July 2002 (Class 181/199); filed 12 July 2000

The figures in this patent are beautifully drawn and unusually complete. The invention itself is a fairly conventional rear loading horn loudspeaker system. Highly damped coupling chamber **34** drives a coiled, slowly expanding air column. A novel feature is the bifurcated mouth, terminating in side panel openings **43** and **44**. This configuration is somehow supposed to, ''...prevent back waves from reflecting back to the driver.''—GLA

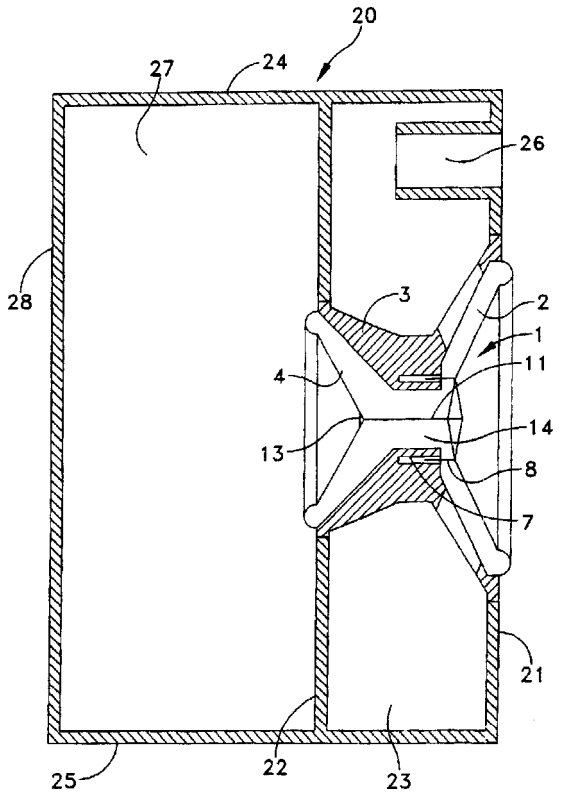


6,431,309

43.38.Ja LOUDSPEAKER SYSTEM

C. Ronald Coffin, Topsfield, Massachusetts
13 August 2002 (Class 181/156); filed 14 April 2000

The inventor's previously patented dual-cone loudspeaker design effectively shrinks an 18-in. woofer into a 15-in. frame. Although sound from the front surfaces of primary cone **1** and auxiliary cone **4** combine to produce front radiation, the rear surfaces of the two cones can be loaded independently. The arrangement shown will work, but not as efficiently as a single-chamber vented box.—GLA

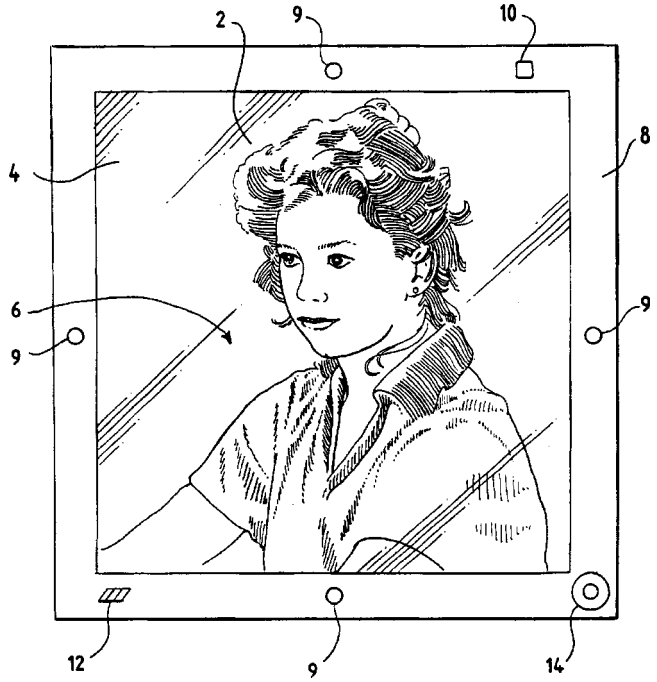


6,393,402

43.38.Md METHOD FOR PRODUCING REMOTELY A PICTURE DISPLAY DEVICE STORING ONE OR MORE ASSOCIATED AUDIO MESSAGES

Alan R. Loudermilk and Wayne D. Jung, assignors to LJ Talk LLC
21 May 2002 (Class 704/272); filed 6 December 2001

Several patent reviews have previously appeared in this journal on how to make a picture frame talk. This one listens, as well. The viewer may hear a particular sound byte associated with a picture by various means, such as touching a point on the frame. By touching another spot, new sounds may be recorded to be added to those already available.—DLR



6,421,446

43.38.Vk APPARATUS FOR CREATING 3D AUDIO IMAGING OVER HEADPHONES USING BINAURAL SYNTHESIS INCLUDING ELEVATION

Terry Cashion and Simon Williams, assignors to QSound Labs, Incorporated
16 July 2002 (Class 381/17); filed 11 December 1998

This patent is a continuation of two earlier patents, the first of which was filed in 1998. The goal is to allow virtual sound sources, as reproduced through headphones, to be located anywhere and moved to anywhere else. The patent text explains that by providing several ''range control blocks'' and ''location control blocks,'' it is possible to work with a fixed number of HRTF filters regardless of the number of audio inputs.—GLA

6,421,447

43.38.Vk METHOD OF GENERATING SURROUND SOUND WITH CHANNELS PROCESSING SEPARATELY

Yu-Ming Chu, assignor to Inno-Tech Company, Limited
16 July 2002 (Class 381/18); filed 30 September 1999

The patent describes yet another method of enhancing two-channel stereo program material to generate an expanded sound field. Instead of

6,424,720

43.38.Lc METHOD AND DEVICE FOR REDUCING MULTI-CHANNEL ACOUSTIC ECHO AND ADAPTING SOUND TO SPACE CONDITIONS

Jean-Philippe Thomas *et al.*, assignors to France Telecom
23 July 2002 (Class 381/66); filed in France 21 May 1997

Persons not directly involved with audio and video teleconferencing may not appreciate the extent to which this technology is used by corporations and other organizations. Several locations may be linked in an electronic conference, with several participants at each location. For optimum speech clarity, each microphone channel should pick up the direct sound from a single talker and reject all other signals, including multiply reflected room sound. Digital signal processing can come surprisingly close to this goal but may become confused when two or more participants at different locations talk simultaneously. Microphones that should be activated may be muted and vice versa. The invention describes an improved algorithm for dealing with this situation.—GLA

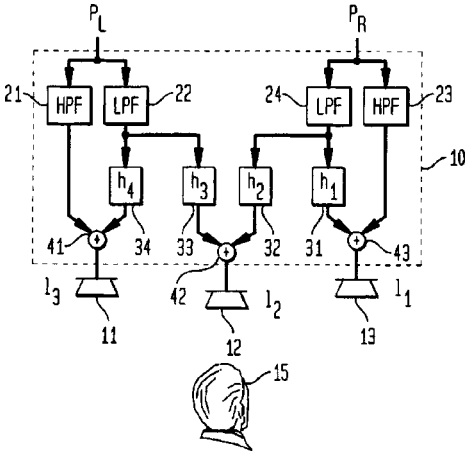
working with sum and difference signals, the left and right signals are processed separately and then subsequently cross coupled.—GLA

6,424,719

43.38.Vk ACOUSTIC CROSSTALK CANCELLATION SYSTEM

Gary W. Elko and Darren B. Ward, assignors to Lucent Technologies Incorporated
23 July 2002 (Class 381/1); filed 29 July 1999

To facilitate the creation of virtual sound sources, audio engineers have spent a lot of time trying to make loudspeakers sound like headphones. The basic problem is that each ear hears both loudspeakers. One way to partly cancel this unwanted interaural crosstalk requires pairs of closely spaced loudspeakers for left and right channels. Other patented methods rely exclusively on electronic processing. For these systems to work effectively, the



listener must remain at one preferred location and not move his or her head. This patent describes an interesting method of enlarging the effective listening area which involves adding a third loudspeaker **12**, set somewhat forward of left and right speakers **11** and **13**. The patent is very well written in terse, technical English accompanied by informative diagrams.—GLA

6,430,293

43.38.Vk RECORDING AND PLAY-BACK TWO-CHANNEL SYSTEM FOR PROVIDING A HOLOPHONIC REPRODUCTION OF SOUNDS

Luca Gubert Finsterle, Cernusco sul Naviglio Milan, Italy
6 August 2002 (Class 381/17); filed in Italy 13 August 1996

The text of this United States patent appears to have been translated from Italian by someone with an excellent command of literary English, but not all that familiar with American technical terms—loudspeakers are called diffusers, for example. Even so, the invention is difficult to understand on its own terms. It appears that signals from two ideally located microphones are recorded directly onto two audio channels, reminiscent of Mercury ''Living Presence'' stereo recordings. Playback requires two conventional front speakers plus two rear speakers. Signals to the rear speakers are essentially the same as the front signals, ''...but having a sound pressure adjusted in a range from −15 dB to +9 dB, preferably from −10 dB to +6 dB.'' If all four speakers are identical and properly adjusted, then a 360 degree reconstruction of the original sound scene can be attained ''...since it will be possible to reduce the perspective distortion occurring in listening from the front diffusers only.''—GLA

6,430,294

43.38.Vk SOUND IMAGE LOCALIZATION METHOD AND APPARATUS, DELAY AMOUNT CONTROL APPARATUS, AND SOUND IMAGE CONTROL APPARATUS WITH USING DELAY AMOUNT CONTROL APPARATUS

Akihiro Fujita *et al.*, assignors to Kabushiki Kaisha Kawai Gakki Seisakusho
6 August 2002 (Class 381/17); filed in Japan 22 October 1996

The patent is concerned with controlling phantom sound sources in a virtual reality system. As the title suggests, the proposed methodology is fairly complicated but is said to reduce computational and memory requirements in comparison with known techniques.—GLA

6,430,535

43.38.Vk METHOD AND DEVICE FOR PROJECTING SOUND SOURCES ONTO LOUDSPEAKERS

Jens Spille and Johannes Böhm, assignors to Thomson Licensing, S.A.
6 August 2002 (Class 704/500); filed in Germany 7 November 1996

The spatial position of each sound source is encoded and recorded along with its acoustic output to create an ''acoustic object.'' During playback the position information is used to create the best possible stereo effect from whatever loudspeaker arrangement is in use.—GLA

6,393,895

43.40.Le METHOD AND APPARATUS FOR CHARACTERIZING MATERIALS BY USING A MECHANICAL RESONATOR

Leonid Matsiev *et al.*, assignors to Symyx Technologies, Incorporated
28 May 2002 (Class 73/24.06); filed 12 August 1998

A mechanical resonator, such as a tuning fork or a piezoelectric element, is immersed in a well that contains a small amount of a liquid that is to be characterized. An input signal is sent to the resonator and swept over a selected frequency range while the resonator's response is measured. The liquid's characteristics are then deduced from that response. Multiple resonators of different types may be used to evaluate a wider range of properties, and some resonator components also may be used to measure the liquid's electrical parameters.—EEU

6,412,586

43.40.Tm TOROIDAL EXHAUST VIBRATION ABSORBER

Gerald W. Askew, assignor to International Truck Intellectual Property Company, L.L.C.
2 July 2002 (Class 180/309); filed 27 May 1999

In order to reduce the vibrations of a heavy truck's vertical exhaust pipe, a dynamic absorber (or ''tuned damper'') is placed around the pipe near its upper end. The absorber consists of a toroidal mass that is attached to the pipe via resilient elements, which essentially consist of vertically or circumferentially oriented leaf springs.—EEU

6,401,933

43.40.Yq DISPLACEABLE ECCENTRIC FOR VIBRATORY SCREEN

Douglas J. Cohen and Mauricio A. Escobar, assignors to Ohio Central Steel Company An Ohio Corporation
11 June 2002 (Class 209/366.5); filed 1 November 2000

This patent pertains to a device that generates vibrations by means of an eccentric mass mounted on a rotating shaft. Here the eccentric mass is arranged so that it can slide radially in response to the centrifugal force that acts on it. Springs keep the mass seated with its center of gravity near the shaft axis until a predetermined rotational speed is reached, thus keeping vibrations small until the machine approaches its operating speed.—EEU

6,404,104

43.40.Yq VIBRATION TYPE ACTUATOR AND VIBRATION TYPE DRIVING APPARATUS

Takashi Maeno *et al.*, assignors to Canon Kabushiki Kaisha
11 June 2002 (Class 310/323.02); filed in Japan 27 November 1997

This actuator for producing motion of a spherical moving member (rotor) about three axes employs only a single vibrating member (stator), unlike prior art devices that used multiple stators. Typical stator embodiments consist of assemblies of piezoelectric elements that may be made to vibrate in different modes by the application of appropriately phased voltages.—EEU

6,405,591

43.40.Yq WHEEL BALANCER USING CONTROLLED LOAD ROLLER FORCE VARIATION

Nicholas J. Colarelli III *et al.*, assignors to Hunter Engineering Company
18 June 2002 (Class 73/462); filed 5 January 2001

This system for balancing automobile wheels incorporates a roller that applies a load to the tire so as to simulate operation of the wheel under road conditions. A controller is used to vary the force applied by the load roller. Data from vibration and force sensors mounted in the axle are fed to a processor and used to determine wheel imbalance, runout, and tire stiffness.—EEU

6,437,568

43.50.Gf LOW NOISE MRI SCANNER

William Alan Edelstein *et al.*, assignors to General Electric Company
20 August 2002 (Class 324/318); filed 2 October 2000

This patent presents a low noise imaging system for producing magnetic resonance images of a subject and for minimizing acoustic noise generated during imaging. The imaging apparatus consists of a magnetic assembly, a gradient coil assembly, and a rf coil assembly. At least one of these assemblies is configured to reduce the generation and transmission of acoustic noise in and about the imaging apparatus. A layer of acoustic absorptive material may be deployed between the conductors and the patient bore tube.—DRR

6,433,693

43.60.Bf APPARATUS AND METHOD FOR BOIL PHASE DETECTION BASED ON ACOUSTIC SIGNAL FEATURES

Harry Kirk Mathews, Jr., assignor to General Electric Company
13 August 2002 (Class 340/584); filed 31 July 2000

Various boil phases (such as presimmer, simmer onset, simmer, and boiling) in utensils on range cook-tops are detected by monitoring the acoustic signals from these utensils. The acoustic signals are conditioned and filtered, and their first derivatives are determined. The results are fed to a feature recognition algorithm.—EEU

6,449,368

43.60.Gk MULTIDIRECTIONAL AUDIO DECODING

Mark Franklin Davis *et al.*, assignors to Dolby Laboratories Licensing Corporation
10 September 2002 (Class 381/1); filed 14 March 1997

This low-mips, crosstalk-cancelling audio network software may be implemented in real time on a personal computer using only a small fraction of the available CPU cycles. The network is particularly useful for rendering surround sound images outside the space between the left and right computer/multimedia loudspeakers. The network includes two signal feedback paths, each having a time delay and frequency-dependent characteristic. The frequency-dependent characteristic represents the smoothed difference in the acoustic attenuation from a transducer to the listener's ear farthest from the transducer and to the listener's ear closer to that transducer. One or more simple digital filters requiring low processing power implement the smoothed difference in the attenuation.—DRR

6,449,371

43.60.Gk PC SURROUND SOUND MIXER

Hock Guan Tan and Edward K. W. Law, assignors to Creative Technology Limited
10 September 2002 (Class 381/119); filed 17 February 1999

The patent relates to a PC surround sound mixing circuit, in particular, a circuit configured to receive multiple audio signals (both from a PC and from other sources), perform a function on the audio signals, and generate a single, multi-channel, surround-sound output signal.—DRR

6,449,215

43.60.Lq THREE-DIMENSIONAL IMAGING SYSTEM FOR SONAR SYSTEM

Richard R. Shell, assignor to the United States of America as represented by the Secretary of the Navy
10 September 2002 (Class 367/7); filed 9 October 2001

A three-dimensional sonar imaging system concept is discussed that processes data from a forward-looking sonar and a side-looking sonar, as well as vehicle position and navigation data, bathymetry data, bottom sediment data, water surface data, and detection/cluster data, to produce a true three-dimensional image of the ensonified region.—WT

6,448,283

43.64.Gz METHODS FOR PREVENTING/TREATING DAMAGE TO SENSORY HAIR CELLS AND COCHLEAR NEURONS

Jukka Ylikoski *et al.*, assignors to Cephalon, Incorporated
10 September 2002 (Class 514/411); filed 24 September 1999

Methods for preventing or treating damage to sensory hair cells and cochlear neurons are disclosed. The methods consist of the administration of an effective amount of a compound of Formula I or Formula II, which are described chemically in extreme detail in the patent. This method is said to provide prevention or treatment of both hearing loss and loss of sense of balance.—DRR

6,397,198

43.66.Lj TOKENLESS BIOMETRIC ELECTRONIC TRANSACTIONS USING AN AUDIO SIGNATURE TO IDENTIFY THE TRANSACTION PROCESSOR

Ned Hoffman *et al.*, assignors to Indivos Corporation
28 May 2002 (Class 705/44); filed 20 July 1999

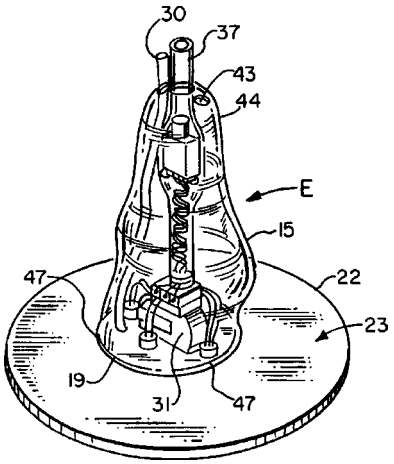
As a part of this biometric identification system, a fragment of sound is generated which will make it easier for the user to uniquely identify another party to a transaction. The sound may be a fragment of speech or music, or it may be generated directly from the digital data using a series of sine- and square-wave generators. The user's perception of this audio signature provides a level of confidence that the digital data is correct much faster than if the user had to directly evaluate the digital data. However, the system is probably not very resistant to deception or mimicry.—DLR

6,438,244

43.66.Ts HEARING AID CONSTRUCTION WITH ELECTRONIC COMPONENTS ENCAPSULATED IN SOFT POLYMERIC BODY

Roger P. Juneau *et al.*, assignors to Softear Technologies
20 August 2002 (Class 381/322); filed 28 October 1998

Materials and method are described for construction of a soft hearing aid that is solid. The body portion of the hearing instrument is constructed of an elastomer having a durometer of 10–35 Shore A. The electronic components are embedded in the solid portion of a soft fill made of silicone or



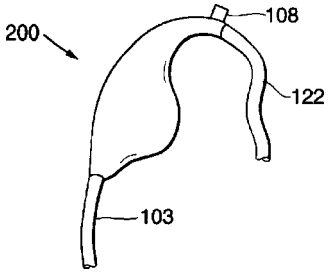
silicone polymer material. The resulting hearing aid is said to be more accurately shaped to the ear since the ear impression is not modified as is done in traditional acrylic shell construction. Additionally, the outer surface of the hearing aid is said to be nonabsorbent and virtually impervious to discoloration, cracking, and ear wax.—DAP

6,438,245

43.66.Ts HEARING AID COMMUNICATIONS EARPIECE

Jon C. Taenzer and William D. Mercer, assignors to ReSound Corporation
20 August 2002 (Class 381/330); filed 2 November 1998

Using cell phones with hearing aids can be problematic due to interference induced from pulsing rf signals via rectification in the hearing aid and acoustic feedback caused by covering the hearing aid microphone inlet. To increase the distance between the rf circuitry and the hearing aid, a detachable earpiece **200** containing a two-way wireless receiver and an acoustic output tube is mounted behind the ear. The earpiece is retained with a break-away clip attached adhesively or via magnetic means to the surface



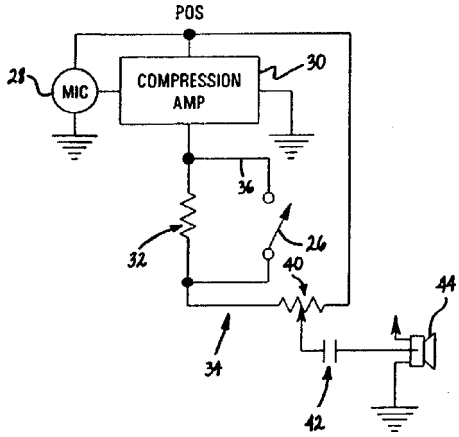
of a hearing aid. The acoustic tube couples the output of the earpiece to the microphone inlet of the hearing aid, which may be an in-the-ear, in-the-canal, or behind-the-ear type. The earpiece may also contain a microphone for picking up and relaying the wearer's voice and environmental sounds back to the two-way communication device. The earpiece may be provided with a transmitter telecoil to couple magnetically to a receiving telecoil in the hearing aid in addition to or in place of the acoustical coupling tube.—DAP

6,442,279

43.66.Ts ACOUSTIC CONDITIONER

David A. Preves *et al.*, assignors to Micro Ear Technology, Incorporated
27 August 2002 (Class 381/72); filed 13 October 2000

For hearing-impaired persons working in noisy environments, earplugs or earmuffs attenuate desired sounds as well as undesired sounds, making communication difficult. A switch **26** enables either normal amplification for a hearing-impaired person or an attenuation mode with some amplification



that also limits to a maximum level the excessive noise levels that can reach the ear. The switch can be controlled manually by the wearer via a toggle switch on the device's housing or automatically in response to the presence of a high long-term average noise level relative to the level for speech or music sounds.—DAP

6,449,372

43.66.Ts METHOD FOR MATCHING HEARING AIDS BINAURALLY

Michael Greminger, assignor to Phonak AG
10 September 2002 (Class 381/314); filed in the European Patent Office 5 January 1999

For ears with symmetrical hearing loss, binaural matching of hearing aid performance is thought to optimize the fitting. Described is a programming system that allows optional simultaneous single adjustment of both left and right hearing aids. To accomplish this, the hearing aids may interact with each other via either a wired or wireless means. Various means for accomplishing this graphically on a computer screen are described.—DAP

6,447,461

43.66.Yw METHOD AND SYSTEM FOR CONDUCTING A HEARING TEST USING A COMPUTER AND HEADPHONES

Steven Anthony Eldon, assignor to Sound ID
10 September 2002 (Class 600/559); filed 7 November 2000

This device constitutes hearing test equipment designed for nonaudiologists (such as musicians, recording studio engineers, audio professionals, etc.) who desire to obtain for themselves accurate information about their hearing at their own recording facility or at home, using computer equipment and headphones. The user-operated, user-calibrated audiological test system includes software, a calibrator, and specified headphones. The frequency response of the headphones is measured. The soundcard of a computer is used to generate audiological test signals, and either the actual output level of those signals or the analysis of the results is compensated by software for the frequency response of the headphones. The calibrator allows an accurate calibration level to be established for the soundcard input. The software provides for using the procedure to perform several audiological tests.—DRR

6,398,730

43.70.Dn NETWORK-BASED METHOD AND APPARATUS FOR THE DETECTION OF THE DYSLEXIA SYNDROME AND RELATED CVS DISORDERS

Harold N. Levinson, Great Neck, New York
4 June 2002 (Class 600/300); filed 14 February 2000

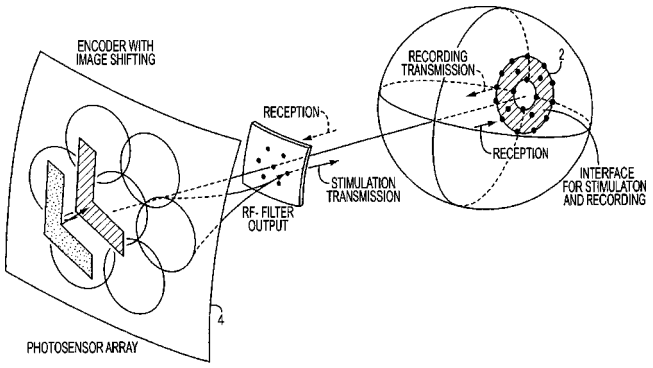
This patent begins with the argument that a large number of language problems, including reading, writing, spelling, and concentration problems, as well as several attention deficit disorders, and even some fears, phobias, and behavioral problems are all related to a known cerebellar-vestibular brain disorder. The patent then presents a system for Internet administration of a test for this disorder, thus making it possible for large numbers of individuals to be tested and making the results of those tests more easily available to the conducting researchers. Some of the tests do require specific visual, auditory, or tactile devices to be installed at the remote computer.—DLR

6,400,989

43.70.Dn ADAPTIVE SENSORY-MOTOR ENCODER FOR VISUAL OR ACOUSTIC PROSTHESIS

Rolf Eckmiller, assignor to Intelligent Implants GmbH
4 June 2002 (Class 607/54); filed in Germany 21 February 1997

Most current systems for implantation of electrode arrays into visual or auditory cortex have depended on an essentially random implantation pattern and on the ability of the cortex to adapt to that random structure to produce meaningful perceptual stimulation. The system described here adds



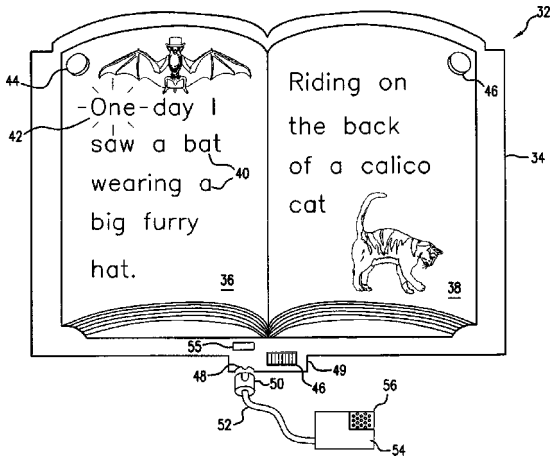
a level of system adaptation, providing a feedback mechanism to allow the patient or the physician to adaptively adjust static and dynamic characteristics of the array excitation according to the patient's perception. Various additional parameters, such as eye or head movement, may be taken into account in the encoding scheme.—DLR

6,405,167

43.70.Ep INTERACTIVE BOOK

Mary Ann Cogliano, Loomis, California
11 June 2002 (Class 704/251); filed 16 July 1999

This device provides reading material for a child, perhaps displayed in a booklike format, and includes speech recognition and playback capabilities. Many individual words in the material would have associated signaling



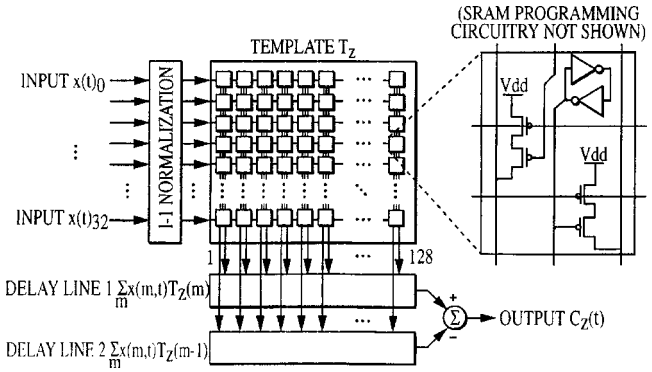
and detection points to call attention to that word or to allow the child to indicate the word to the system. The word could be pronounced for the child or the child's pronunciation could be evaluated by the system.—DLR

6,389,377

43.72.Ar METHODS AND APPARATUS FOR ACOUSTIC TRANSIENT PROCESSING

Fernando J. Pineda *et al.*, assignors to The Johns Hopkins University
14 May 2002 (Class 703/4); filed 1 December 1998

This mixed-signal (analog and digital) integrated circuit design is intended for detecting transient sounds, up to 100 ms in duration, in a model of the human ear. The input to the transient detector is a bank of filters, in this case, 31 bands, log-spaced from 100 to 6000 Hz. By several steps of normalization and differencing, each band signal is reduced to a binary



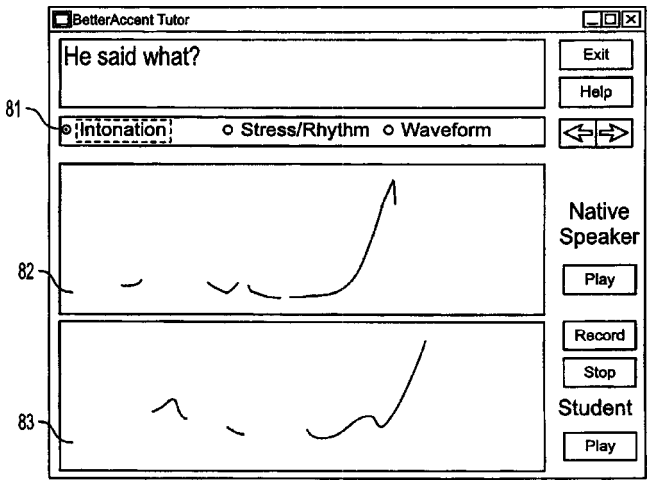
level. This allows simple XOR detection, avoiding either digital or analog multiplication. By concentrating on relatively short transients, the device avoids a need for time adjustments. There is a brief discussion of classification algorithms usable with the resulting detected time-frequency matrix data.—DLR

6,397,185

43.72.Ar LANGUAGE INDEPENDENT SUPRASEGMENTAL PRONUNCIATION TUTORING SYSTEM AND METHODS

Julia Komissarchik and Edward Komissarchik, assignors to Betteraccent, LLC
28 May 2002 (Class 704/270); filed 29 March 1999

This software package for helping a person learn the proper pronunciation of a foreign language provides side-by-side comparisons of the source materials and the user's productions. Any of a number of speech



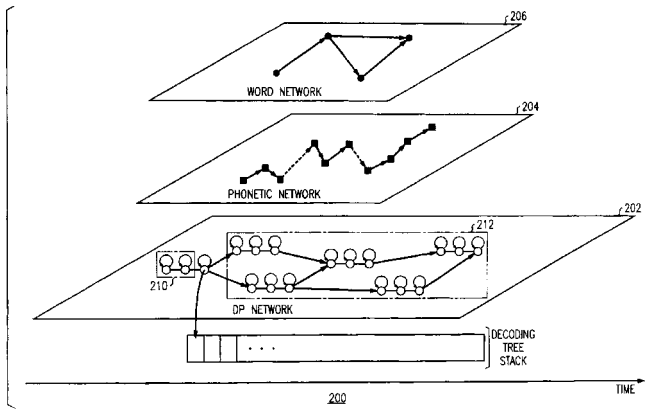
parameters may be selected for display. An algorithm then examines the phonetic and syllabic patterns to locate those aspects of the signal most likely to represent significant problem areas in the student's pronunciation.—DLR

6,442,520

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

Eric Rolse Buhrke and Wu Chou, assignors to Agere Systems Guardian Corporation
27 August 2002 (Class 704/255); filed 8 November 1999

A continuous speech recognizer is proposed having three dynamically expanded networks, each with self-adjusting capability. Speech data is separated into frames and inputted into a dynamic programming network. Nodes are built by the dynamic programming network that represent likelihood scores of predefined models corresponding to the speech data. A phone (speech) expanding network provides phone rules that control which nodes of the dynamic programming network can be connected by arcs to other nodes. The speech framer output is also connected to a word network process for storing and applying language rules. The asynchronous word network operates in parallel with the other two networks to provide word rules that control which portions of the phone network correspond and do not correspond to recognizable words.—DAP



6,405,166

43.72.Fx MULTIMEDIA SEARCH APPARATUS AND METHOD FOR SEARCHING MULTIMEDIA CONTENT USING SPEAKER DETECTION BY AUDIO DATA

Qian Huang *et al.*, assignors to AT&T Corporation
11 June 2002 (Class 704/246); filed 15 October 2001

This is a system for searching through multimedia data, such as a videotape library, for a particular speaker's voice. A large database would contain speaker-specific information, such as Markov models trained for specific, perhaps well-known, individuals. A user could then request a search by giving the target individual's name. The system would locate relevant materials in the large database.—DLR

6,385,584

43.72.Ja PROVIDING AUTOMATED VOICE RESPONSES WITH VARIABLE USER PROMPTING

Alexander I. McAllister and James E. Curry, assignors to Verizon Services Corporation
7 May 2002 (Class 704/275); filed 30 April 1999

This speech playback system, intended for use in a telephone dialing response system, attempts to provide a more natural user interface by storing a variety of alternate phrases to be used in response to a particular situation. One of the various alternate phrases, all intended to be semantically equivalent, is selected randomly when the corresponding situation arises. A moment's reflection would reveal that such alternates are rarely in fact semantically equivalent. If anything, this kind of solution, rather than making the machine seem more human, probably makes it seem less so.—DLR

6,393,400

43.72.Ja INTELLIGENT OPTICAL DISK WITH SPEECH SYNTHESIZING CAPABILITIES

Takashi Shigetomi *et al.*, assignors to Kabushiki Kaisha Optrom
21 May 2002 (Class 704/258); filed in Japan 18 June 1997

As computer disk drive speeds and capacities increase, we expect to see speech processing systems make use of those increases. This patent describes the straightforward application of existing speech playback and synthesis techniques to a large-capacity disk storage system. The novel feature, if it is that, is that the speech processing logic is closely incorporated into the disk drive circuitry, for maximum use of the drive's capabilities.—DLR

6,405,169

43.72.Ja SPEECH SYNTHESIS APPARATUS

Reishi Kondo and Yukio Mitome, assignors to NEC Corporation
11 June 2002 (Class 704/258); filed in Japan 5 June 1998

This speech synthesis system begins with separate subsystems for generating phonetic and prosodic aspects of the speech signal. The signal is then reevaluated to reconcile discrepancies between these two aspects of the signal. In the classical case of interactions between these two signal aspects, the pitch and duration details are modified on the basis of the local phonetic structure. Although the patent suggests that modifications can act in the opposite direction as well, no examples of such are provided.—DLR

6,393,444

43.72.Kb PHONETIC SPELL CHECKER

Stephen Graham Copinger Lawrence, assignor to International Business Machines Corporation
21 May 2002 (Class 707/533); filed in the United Kingdom 22 October 1998

This spell checker converts the input word into a phonetic representation, does a weighted phonetic lookup, and then converts the result back to the spelled form. A 15-year-old paper on phonemes and their spelling is cited as the basis for the text-to-phonetic conversion. The result of this process is a sequence of ''clusters,'' each of which is the longest sequence of letters found to have a specific phonetic correlate. Various steps are described by which the clusters are assigned weightings for each of the possible pronunciations. A set of rules then guides the process of conversion back to the spelled form. Some of the steps in the system can be trained with plain, unmarked text, such as books available via the Internet, rather than requiring a phonetic dictionary.—DLR

6,404,859

43.72.Kb VOICE ENABLED SYSTEM FOR REMOTE ACCESS OF INFORMATION

Moh'd Abdel-Hamid Hasan, assignor to Lockheed Martin Corporation
11 June 2002 (Class 379/88.04); filed 6 May 1999

This system for voice interaction with a computer provides security features allowing certain limited or restricted accesses to sensitive information by voice in a closed, private computer network. The system includes both speaker verification methods as well as more traditional data access protocols.—DLR

6,404,860

43.72.Kb SYSTEM AND METHOD FOR INTERNET CALL MANAGEMENT WITH TEXT-TO-SPEECH MESSAGING

Theodore E. Casellini, assignor to SS8 Networks, Incorporated
11 June 2002 (Class 379/88.17); filed 9 February 2000

The patent describes a series of protocols for combining voice and text messages smoothly in a system for Internet telephone communication. Both notification to the subscriber and feedback to the caller are converted to an appropriate text or voice form depending on preferences and current conditions at each user's unit.—DLR

6,404,872

43.72.Kb METHOD AND APPARATUS FOR ALTERING A SPEECH SIGNAL DURING A TELEPHONE CALL

Randy G. Goldberg *et al.*, assignors to AT&T Corporation
11 June 2002 (Class 379/201.11); filed 25 September 1997

The system, as described in this patent, would add an extra processing layer to a communication network system, such as a cell phone linkage. The new layer would be programmable in some reasonable way by a caller so as to disguise the caller's voice or to alter the caller's signal in any of several ways. Thus, the caller could alter the background so as to pretend that he or she is at a wild party or, perhaps, not at the wild party.—DLR

6,438,521

43.72.Ne SPEECH RECOGNITION METHOD AND APPARATUS AND COMPUTER-READABLE MEMORY

Masayuki Yamada *et al.*, assignors to Canon Kabushiki Kaisha
20 August 2002 (Class 704/254); filed in Japan 17 September 1998

In single-syllable speech recognition systems, a silent period between syllables tends to be recognized as the end of an utterance. The described system determines whether speech separately uttered as a single syllable is included in the input speech by comparing a first score calculated by comparing an arbitrary syllable sequence with the input speech to a second score calculated by a speech recognition result of the input speech. The speech recognition result is output when the second score is larger than the first score. Alternatively, on the basis of the determination result, the system may prompt the user to input speech again by speaking continuously without any pause.—DAP

6,449,594

43.72.Ne METHOD OF MODEL ADAPTATION FOR NOISY SPEECH RECOGNITION BY TRANSFORMATION BETWEEN CEPSTRAL AND LINEAR SPECTRAL DOMAINS

Tai-Hwei Hwang and Hsiao-Chuan Wang, assignors to Industrial Technology Research Institute
10 September 2002 (Class 704/233); filed in Taiwan, Province of China 7 April 2000

Computation requirements are high for adjusting probability density functions when using a parallel modal combination method of combining statistical data or of speech and noise in the linear spectral domain. Past approaches have tried to reduce processing time by adapting mean vectors and ignoring variances. This has resulted in reduced accuracy for the recognition of speech in noise. In the method proposed here, transforming statistical data of speech and noise from the cepstral domain to the linear spectral

domain reduces processing time while maintaining high recognition accuracy. Cepstral mean noise and speech vectors are first transformed into the linear spectral domain. The linear vectors are combined to obtain a linear noisy speech vector which is then transformed back to the cepstral domain. Finally, the cepstral covariance speech and noise matrices are multiplied by scaling factors and the resulting scaled cepstral covariance matrices are combined to determine the cepstral covariance matrix of adapted noisy speech.—DAP

6,393,397

43.72.Pf COHORT MODEL SELECTION APPARATUS AND METHOD

Ho Chuen Choi *et al.*, assignors to Motorola, Incorporated
21 May 2002 (Class 704/250); filed in Australia 17 June 1998

The cohort model of speaker identification refers to the method of searching through many individual speaker models for those which are most similar to a new registering speaker. The model for the new speaker is stored along with references to the set of similar ''cohort'' models. During speaker verification, the candidate's voice must be closest to the claimed speaker's patterns and also closer to those patterns than to any of the cohort models. The patent deals with efficient methods of searching the speaker model database to locate usable cohort models for a new registrant and describes various details of the feature set and the distance measures which affect such searches.—DLR

6,401,063

43.72.Pf METHOD AND APPARATUS FOR USE IN SPEAKER VERIFICATION

Matthieu Hébert and Stephen D. Peters, assignors to Nortel Networks Limited
4 June 2002 (Class 704/234); filed 9 November 1999

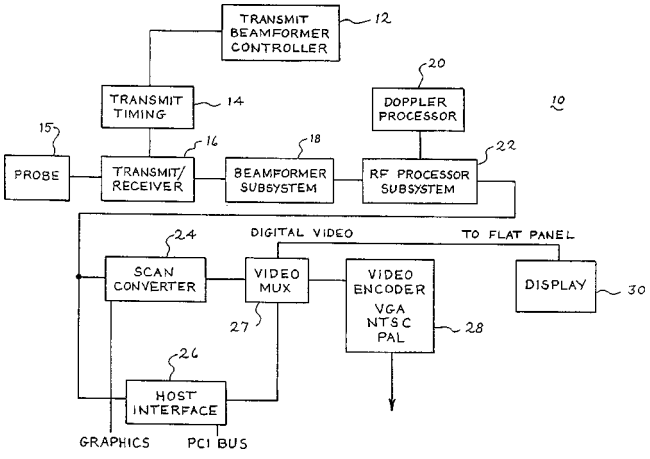
This system for speaker verification processes the speech of a new registrant in two ways using a speaker-independent reference model to produce both speaker-dependent and speaker-independent patterns. These patterns are then further processed to derive a biased normalizing template, which forms the basis for the speaker verification processing.—DLR

6,436,039

43.80.Qf MEDICAL DIAGNOSTIC ULTRASOUND SYSTEM AND METHOD

Anthony P. Lannutti *et al.*, assignors to Ecton, Incorporated
20 August 2002 (Class 600/437); filed 14 September 1999

In this medical diagnostic ultrasound imaging system, two or more components of an image are separately stored, allowing a composite image to be constructed from selected components. While other conventional ultrasound systems use a frame grabber to perform a ''what-you-see-is-what-you-get'' technique of image capture, this system provides selective construction of a composite image, thus facilitating image presentation, review, and post-processing functions.—DRR



6,436,047

43.80.Qf APERTURE CONFIGURATIONS FOR MEDICAL DIAGNOSTIC ULTRASOUND

Bhaskar S. Ramamurthy *et al.*, assignors to Acuson Corporation
20 August 2002 (Class 600/447); filed 27 November 2000

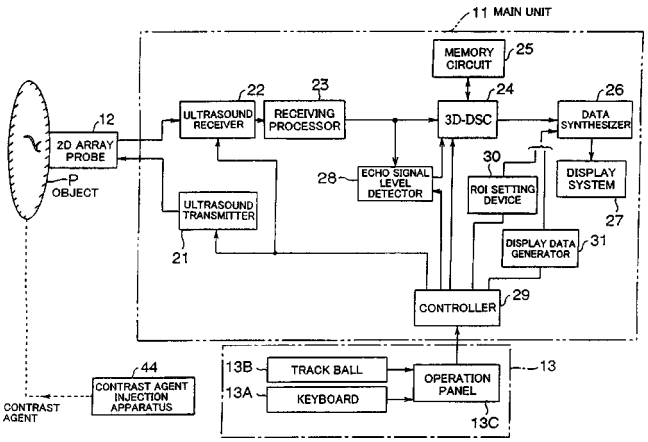
The patent describes a method for obtaining ultrasound data using transducer element switching. The switching is applied to change the effective distances between elements for at least one transmission or reception of acoustic energy. The resulting reflected echo signals are processed to yield ultrasound data. In the preferred embodiment, aperture techniques differ for the transmit and receive phases. The various techniques include synthetic aperture, shorting elements, skipping elements, sliding apertures, and combinations thereof.—DRR

6,436,049

43.80.Qf THREE-DIMENSIONAL ULTRASOUND DIAGNOSIS BASED ON CONTRAST ECHO TECHNIQUE

Naohisa Kamiyama and Yoichi Ogasawara, assignors to Kabushiki Kaisha Toshiba
20 August 2002 (Class 600/458); filed in Japan 31 May 1999

During ultrasound diagnosis, a contrast echo technique for three-dimensional scanning is applied with a contrast agent injected into the object. Before and after injecting the contrast agent, a blood vessel is continuously and securely targeted and depicted in displayed images. To accomplish this, the apparatus acquires an echo signal by scanning a three-dimensional region with an ultrasonic beam, produces the three-dimensional image data, specifies (e.g., from the echo signal) a time at which injection of the contrast



agent into the region undergoing scanning is started, and switches over at a specified time into display states based on the three-dimensional image data. For example, the switchovers can be executed so that the data produced by projecting the three-dimensional image data on a minimum intensity projection technique are displayed before injection of the contrast agent occurs, while data produced by projecting them on a maximum intensity projection technique are displayed after the start of the injection of the contrast agent.—DRR

6,436,057

43.80.Qf METHOD AND APPARATUS FOR COUGH SOUND ANALYSIS

William T. Goldsmith *et al.*, assignors to the United States of America as represented by the Department of Health and Human Services, Centers for Disease Control and Prevention
20 August 2002 (Class 600/586); filed 21 April 2000

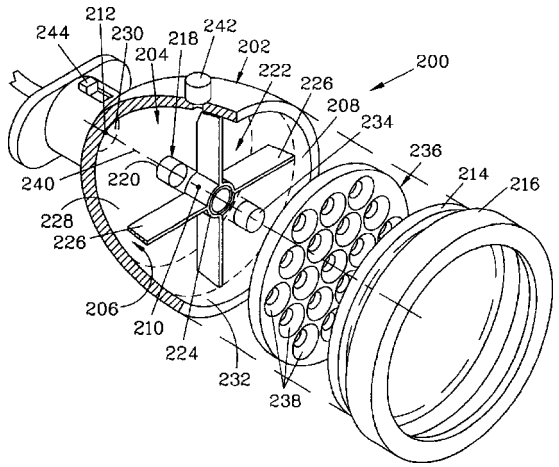
This patent entails a method for recording cough sounds for the purpose of diagnosing pulmonary disorders and diseases. The method applies signal analysis techniques to extract quantitative information from the recorded cough sound pressure waves. An embodiment includes a mouthpiece connected to the proximal end of a tube. A microphone is attached to the tube between the distal and proximal ends for recording sound pressure waves. A cough sound index can be calculated for diagnostic purposes.—DRR

6,438,238

43.80.Qf STETHOSCOPE

Thomas F. Callahan, Grantham, New Hampshire
20 August 2002 (Class 381/67); filed 14 July 2000

The housing of the chest piece for this stethoscope has a cavity with an interior paraboloidal surface **206**. The focal point **210** is inside a support frame in the cavity, which holds transducer **218** positioned at this focal point. The parabolic shape of the cavity surface acts to reflect sound waves entering normal to the front opening back to transducer, while rejecting



other sound waves. A mechanical wave guide with a series of parallel passages extends across the opening. The cavity is preferably filled with a liquid or gel having the acoustical impedance of human flesh. To reduce noise from motion of the chest piece across the skin, the stethoscope can be activated by a pressure switch responsive to fluid pressure inside the cavity.—DRR

6,440,082

43.80.Qf METHOD AND APPARATUS FOR USING HEART SOUNDS TO DETERMINE THE PRESENCE OF A PULSE

Tae H. Joo *et al.*, assignors to Medtronic Physio-Control Manufacturing Corporation
27 August 2002 (Class 600/528); filed 30 September 1999

This apparatus determines the presence of a cardiac pulse in a patient by evaluating a physiological signal, principally the presence of characteristic heart sounds as determined by analysis of phonocardiogram (PCG) data. The analysis may include evaluation of temporal or spectral energy in the PCG data. Electrocardiogram data may also be used to detect the presence of a cardiac pulse in the patient. An automated external defibrillator reports to the operator whether a cardiac pulse is present or not.—DRR

6,443,899

43.80.Qf ULTRASONIC PROCESSES AND CIRCUITS FOR PERFORMING THEM

Volkmar Uhlendorf *et al.*, assignors to Schering Aktiengesellschaft
3 September 2002 (Class 600/458); filed in Germany 1 September 1988

This patent represents an addendum to United States patents 6,221,017, 6,064,628, 6,034,922, and 5,482,316 and covers a process for selective graphic representation and/or evaluation of objects which respond nonlinearly to sonic energy, for example, biological organs and tissues. An examination area is acoustically irradiated, nonlinear oscillations are produced in the targeted area, and an ultrasonic converter evaluates the ensuing signal. A circuit for carrying out the above process is also disclosed.—DRR

6,443,907

43.80.Qf ACOUSTIC DETECTION OF RESPIRATORY CONDITIONS

Hussein A. Mansy and Richard H. Sandler, assignors to Biomedical Acoustic Research, Incorporated
3 September 2002 (Class 600/529); filed 6 October 2000

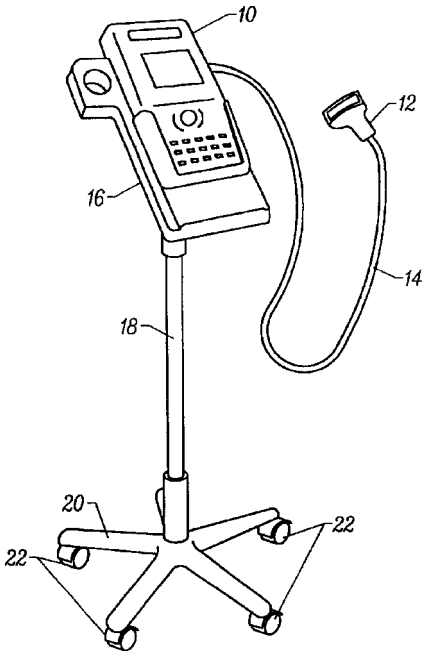
Diagnostic techniques are provided to detect a patient's respiratory condition. The techniques make use of acoustic characteristics of the patient's lungs and chest to determine if a respiratory condition is present. The acoustic generation and transmission characteristics in the lungs and chest are compared with reference acoustic characteristics and/or predetermined threshold values to determine if an abnormal respiratory state exists. The techniques can be applied, for example, to detect the presence of a gas cavity between the lungs and chest wall, which is symptomatic of a pneumothorax condition, or to detect a relative difference between the acoustic transmission characteristics from a patient's trachea to the left and right lungs, symptomatic of an ET tube blocking (or partially blocking) one of the patient's bronchi. The sound waves or vibration may be initiated by a speaker that emits sound waves into the patient's mouth or trachea. Indigenous sounds or percussive inputs into the chest wall may also be used. In processing the sound waves, the energy ratios are calculated using high- and low-frequency band energy, signal time delays, and/or dominant frequencies and the calculated values are compared with predetermined thresholds to indicate the respiratory state of the patient.—DRR

6,447,451

43.80.Qf MOBILE ULTRASOUND DIAGNOSTIC INSTRUMENT AND DOCKING STAND

Gregory Wing *et al.*, assignors to SonoSite, Incorporated
10 September 2002 (Class 600/437); filed 3 May 2000

This device is a mobile ultrasound diagnostic instrument that includes a self-powered ultrasound console incorporating the electronics for driving a transducer array and processing reflected ultrasound waves and a visual



display for the processed ultrasound waves. A docking stand is provided for the console which includes a variable-height sleeve to receive the instrument console. The sleeve can serve as a conduit for signal and power connectors and is mounted on a wheeled dolly.—DRR

6,447,455

43.80.Qf ULTRASOUND DIAGNOSTIC APPARATUS AND METHOD FOR MEASURING BLOOD FLOW VELOCITY USING DOPPLER EFFECT

Ji Hoon Bang and Cheol An Kim, assignors to Medison Company, Limited
10 September 2002 (Class 600/454); filed in the Republic of Korea 8 July 2000

The objective of this device is to determine whether or not aliasing occurs in sample data while measuring blood flow velocity in a human body. A diagnostic method is provided for updating a pulse repetition frequency (PRF) so as to prevent the sample data aliasing. Using the Doppler effect, sample data is generated by transmitting an ultrasound signal into a human body and sampling a reflected signal. Frequency distribution data generated by processing the sample data forms the basis for detecting blood flow velocity. The processing detects whether or not aliasing has occurred in the frequency distribution data. When aliasing is detected, the PRF is updated in order to prevent it.—DRR

6,436,060

43.80.Sh SUBMERSIBLE SYSTEM FOR ULTRASONIC TREATMENT

Roger J. Talish, assignor to Exogen, Incorporated
20 August 2002 (Class 601/2); filed 18 October 1999

This is a system used for therapeutically treating injuries using ultrasound. It includes an ergonomically constructed ultrasonic transducer treatment head module and a main operating unit. The head module is positioned adjacent to the injured area and excited for a predetermined period of time. The system also includes a bathtub insert that envelops a portion of the patient's body and a way of positioning and holding the treatment head module with respect to the patient's body.—DRR

6,436,061

43.80.Sh ULTRASOUND TREATMENT OF VARICOSE VEINS

Peter D. Costantino, Armonk, New York
20 August 2002 (Class 601/2); filed 29 December 1999

The patent describes a method for noninvasive treatment of varicose veins. The vibrational energy from one or more ultrasonic sources is focused beneath the skin to cause localized heating and tissue disruption and/or destruction in and around the varicosed segment of the vein, particularly the venous endothelium. The localized tissue destruction is followed by localized fibrosis in the region of the dilated portion of the varicosed vessel. If a sufficiently high dosage of ultrasound energy is delivered to the vein, the varicosed portion of the vessel will be obliterated, yielding a therapeutic result similar to that obtained by injection sclerotherapy or vein stripping. Also, focused ultrasound may be used to induce fibrosis in the tissue overlying the varicosed vein thereby providing a relatively inelastic barrier that resists unsightly deformation of the skin overlying the varicose vein.—DRR

6,436,114

43.80.Sh APPARATUS FOR TREATING BODY TISSUE WITH ULTRASOUND

Pavel Novak *et al.*, assignors to Storz Endoskop GmbH
20 August 2002 (Class 606/169); filed in Germany 18 February 1998

The apparatus is a surgical device that not only suctions out and discharges coagulated and fluidized cell and tissue components (and optionally injects a flushing fluid into the operation area and removes it again), but also includes a cutting blade and two grasping jaws in the region of the distal end of a hollow probe.—DRR

6,439,236

43.80.Sh METHODS FOR INDUCING ATRIAL AND VENTRICULAR RHYTHMS USING ULTRASOUND AND MICROBUBBLES

Thomas R. Porter and Feng Xie, assignors to The Board of Regents of the University of Nebraska
27 August 2002 (Class 128/898); filed 25 October 1999

A cardiac rhythm therapy that is noninvasive, supposedly painless, and nontoxic involves the combination of low-frequency transthoracic ultrasound and a pharmaceutical composition. The composition contains microbubbles of a diameter of about 0.1 to 10 microns, filled with an insoluble gas such as fluorocarbon gas, helium, or sulfur hexafluoride. The gas is encapsulated in a protein-coated shell. The device uses agents and methods normally used in ultrasound imaging and as such provides a means for visualization of the heart as the rhythm is induced. It was found (quite unexpectedly) that the insoluble gas microbubbles in combination with low-frequency ultrasound induce or pace atrial and ventricular rhythms in animals and induction can be manipulated by directing the ultrasound to either the atrium or ventricle. This rhythm strategy can be used in a number of ways.—DRR

6,440,077

43.80.Sh APPARATUS AND METHOD FOR THE INTRAVASCULAR ULTRASOUND-GUIDED PLACEMENT OF A VENA CAVA FILTER

Matthew T. Jung *et al.*, all of Louisville, Kentucky
27 August 2002 (Class 600/467); filed 9 May 2000

This is an apparatus for the intravascular placement of a vena cava filter for treatment of deep vein thrombosis. The apparatus includes a

sheathed intravascular ultrasound catheter with an ultrasonic imaging element and the vena cava filter that is to be deployed on a guide wire. The outer sheath, catheter, and guide wire share a common central axis. The catheter is enclosed by and is movable relative to the sheath and the guide wire is enclosed by and is movable relative to the catheter. In the stored position, the filter is secured between the sheath and the catheter. When the device is introduced into a vein, the catheter provides real-time imaging of the vein for identifying the location appropriate for filter placement. Once such a location has been identified, the outer sheath is drawn back relative to the catheter, exposing the legs of the filter, thereby permitting the legs to spring free and attach themselves to the wall of the vein.—DRR

6,425,872

43.80.Vj SYSTEM FOR MEASURING PHYSICAL PARAMETER UTILIZING VIBRATION TRANSMISSION

Hisashi Hagiwara *et al.*, assignors to Matsushita Electric Industrial Company, Limited
30 July 2002 (Class 600/485); filed in Japan 20 October 1999

The transmission and detection of acoustic waves through human tissue can be used for various noninvasive medical diagnostic procedures. Many of these rely on computer analysis to achieve a degree of precision that would have seemed miraculous only a few years ago. This patent, for example, describes a method of taking a patient's blood pressure without physical contact.—GLA

6,436,041

43.80.Vj MEDICAL ULTRASONIC IMAGING METHOD WITH IMPROVED ULTRASONIC CONTRAST AGENT SPECIFICITY

Patrick J. Phillips and Ismayil M. Guracar, assignors to Acuson Corporation
20 August 2002 (Class 600/437); filed 22 December 2000

Echo signals received from pulses of a multiple-pulse sequence are acquired and combined using weights that are selected to suppress energy at either the fundamental frequency or a desired harmonic of the fundamental frequency. In this method, an additional pulse or pulses are transmitted or modifications are made to a transmit pulse sequence across different scan lines. The method can use any multiple-pulse sequence that varies the amplitude or phase between transmit pulses to improve agent-to-tissue specificity and image quality when ultrasonic contrast agents are present.—RCW

6,436,044

43.80.Vj SYSTEM AND METHOD FOR ADAPTIVE BEAMFORMER APODIZATION

Hong Wang, assignor to Siemens Medical Solutions USA, Incorporated
20 August 2002 (Class 600/443); filed 14 February 2000

Two parallel beamformers, each followed by an amplitude detector, use different apodization functions. The differently apodized signals are compared and then combined to produce a signal for a single beam that results in a desired lateral and contrast resolution simultaneously.—RCW

6,436,046

43.80.Vj DIAGNOSTIC ULTRASOUND IMAGING METHOD AND SYSTEM WITH IMPROVED FRAME RATE

David J. Napolitano *et al.*, assignors to Acuson Corporation
20 August 2002 (Class 600/447); filed 27 October 2000

Receive beams that alternate in type are formed from spatially distinct transmit beams. The types of beams differ in at least one scan parameter other than transmit and receive line geometry. The parameters used may be transmit phase, transmit or receive aperture, system frequency, transmit focus, phase angle, transmit code, or gain. The receive beams are combined to reduce the frame rate penalty normally associated with techniques such as phase inversion, synthetic aperture, and synthetic focusing.—RCW

6,438,258

43.80.Vj ULTRASOUND IMAGE PROCESSING EMPLOYING NON-LINEAR TISSUE RESPONSE BACKSCATTER SIGNALS

George A. Brock-Fisher and Thomas J. Hunt, assignors to Koninklijke Philips Electronics N.V.
20 August 2002 (Class 382/128); filed 23 January 1998

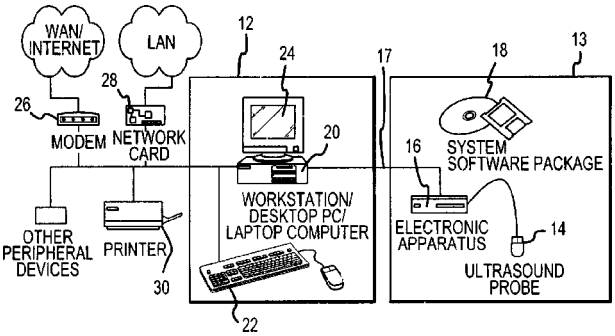
Images are enhanced through the use of both harmonic and fundamental signals. Samples of backscattered signals are obtained in a range of frequencies about a transmitted fundamental frequency and a range of frequencies about a harmonic of the fundamental. Control values are then derived from either the fundamental or harmonic signals and used to control processing of other signals. The control values are also used to alter, e.g., by variation of gain or filter frequency and bandwidth, the processing of the fundamental signals.—RCW

6,440,071

43.80.Vj PERIPHERAL ULTRASOUND IMAGING SYSTEM

Michael H. Slayton *et al.*, assignors to Guided Therapy Systems, Incorporated
27 August 2002 (Class 600/437); filed 18 October 1999

This system **13** connects to a personal computer or a computer network and permits real-time ultrasonic imaging without additional modules or the need to modify or replace the existing central processor.—RCW



6,440,073

43.80.Vj ULTRASONIC DIAGNOSTIC IMAGING SYSTEM WITH AUTOMATIC RESTART AND RESPONSE

Andrew L. Robinson *et al.*, assignors to Koninklijke Philips Electronics N.V.
27 August 2002 (Class 600/437); filed 24 October 2000

This imaging system will automatically turn on at predetermined times and condition itself for a specific scanning procedure. The system can also respond to remote queries when not in service. This permits the system to be off at night, to respond to external queries during that time, and be ready for use at a desired time the next day, all without operator intervention.—RCW

6,440,075

43.80.Vj ULTRASONIC DIAGNOSTIC IMAGING OF NONLINEARLY INTERMODULATED AND HARMONIC FREQUENCY COMPONENTS

Michalakis Averkiou, assignor to Koninklijke Philips Electronics N.V.
27 August 2002 (Class 600/443); filed 6 August 2001

Nonlinearly produced echo signals that contain both harmonic and intermodulation products are separated from the fundamental frequency components of the signals by pulse inversion processing. The use of both nonlinear components permits improvement of the signal-to-noise ratio in ultrasonic images. The two components can be blended or used in different regions of an image to offset the effects of depth-dependent attenuation.—RCW

6,443,894

43.80.Vj MEDICAL DIAGNOSTIC ULTRASOUND SYSTEM AND METHOD FOR MAPPING SURFACE DATA FOR THREE DIMENSIONAL IMAGING

Thilaka S. Sumanaweera *et al.*, assignors to Acuson Corporation
3 September 2002 (Class 600/443); filed 29 September 1999

The boundary of a structure is determined from one type of ultrasonic echo data such as Doppler energy. Using the boundary, another type of data, such as b-mode, that represents the boundary or area adjacent to the boundary is extracted or identified. The second type of data, e.g., b-mode, is then rendered as a function of the boundary, such as by texture mapping, onto or adjacent to the boundary. The texture mapped data may provide information by itself or be used to select data for volume rendering.—RCW

6,443,895

43.80.Vj WAVELET DEPULSING OF ULTRASOUND ECHO SEQUENCES

Dan Adam and Oleg Michailovich, assignors to Technion Research and Development Foundation Limited
3 September 2002 (Class 600/443); filed 10 September 2001

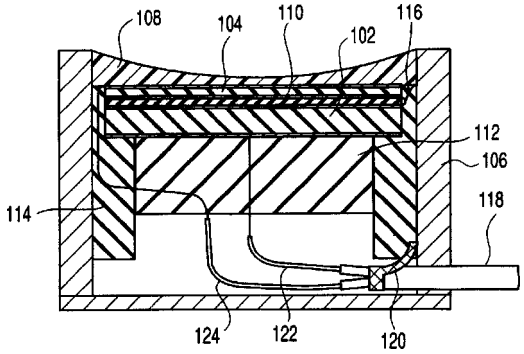
An echo sequence from an object is acquired and a log spectrum of at least a portion of the echo sequence is computed. A low-resolution, shift-invariant wavelet projection of the log spectrum is used as an estimate of a log spectrum of the point-spread function. The frequency-domain phase of the point-spread function is also estimated. Scatterer reflectivity is deconvolved from the echo sequence using the estimated point-spread function.—RCW

6,443,900

43.80.Vj ULTRASONIC WAVE TRANSDUCER SYSTEM AND ULTRASONIC WAVE TRANSDUCER

Hideo Adachi and Katsuhiro Wakabayashi, assignors to Olympus Optical Company, Limited
3 September 2002 (Class 600/458); filed in Japan 15 March 2000

This system consists of an ultrasonic transducer that transmits a signal at a fundamental ultrasound frequency, another transducer to receive harmonic signals having a center frequency that is an integer multiple of the



fundamental, and a controller that provides a drive signal and desired terminations.—RCW

Comment on "Investigation of sound waves generated by the Hall effect in electrolytes" [J. Acoust. Soc. Am. 111(5), 2087–2096 (2002)] (L)

Andrzej B. Dobrucki[a)]
Institute of Telecommunications and Acoustics, Wroclaw University of Technology, Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland

(Received 23 September 2002; accepted for publication 16 December 2002)

An improved model of magnetohydrodynamic transducer is presented. According to this model, the equations for direct and inverse transduction effect have the same form and they fulfill the reciprocity principle. The general form of equation for transduction coefficient in magnetohydrodynamic transducers is given
[DOI: 10.1121/1.1543589]

PACS numbers: 43.38.Ar, 43.38.Dv, 43.35.Rw [AJZ]

The curve for the case of rigid end of the tube in Fig. 5(a) of the cited paper[1] is incorrect. The transfer function of a magnetohydrodynamic (MHD) transmitter has the form (the same symbols as in Ref. 1 are used)

$$G_L(ks)=j\,\frac{1-\cos(ks)}{ks}, \tag{1}$$

and its absolute value is presented in Fig. 1.

This transfer function is identical to the transfer function for receiver with the same rigid boundary condition at $z=0$, except for the factor 2. Transfer functions for all considered cases are the same for loudspeaker as well as for receiver, except for the factor 2. Then, Fig. 5(b) is not necessary.

The unhomogeneous wave equation (16) in the cited paper produces two waves: first traveling in positive-z direction and the second one in negative-z direction. For the cases of rigid and free boundary conditions at $z=0$ the second wave reflects from the boundary and interferes with the first wave. For matched impedance at $z=0$ the second wave is absorbed by this impedance. The identical quantitative reciprocal effect is caused by the exciting wave corresponding to the first generated wave, i.e., traveling in negative-z direction with amplitude of velocity equal to *half* the amplitude of velocity of the entire generated wave. The second wave (reflected for rigid and free boundary conditions and absorbed for matched impedance) appears also in the receiver. It also influences reciprocal effect. Then, in Eq. (35) of the paper in question, the amplitude of the velocity of the wave causing the reciprocal electric effect should be equal to $\xi_{,to}/2$ rather than $\xi_{,to}$. The factor equal to 2 vanishes for all receiver transfer functions, and receiver transfer functions are equal to corresponding transmitter transfer functions

$$G_M(ks)=G_L(ks)=G(ks), \tag{2}$$

where $G(ks)$ for rigid boundary condition is given by Eq. (1)

$$G(ks)=\frac{\sin(ks)}{ks} \tag{3}$$

for free boundary condition, and

$$G(ks)=\frac{\sin(ks/2)}{ks/2}\cdot\frac{e^{j(ks/2)}}{2} \tag{4}$$

for matched impedance at $z=0$.

The pair of reciprocal equations for an electroacoustic transducer of magnetic type has the form[2]

$$p=TI,\quad V=TU, \tag{5}$$

where p=acoustic pressure, U=volume velocity, I=electric current, V=voltage, T=transduction coefficient. The volume velocity is expressed by the equation

$$U=\xi_{,to}wh, \tag{6}$$

and the voltage of the MHD receiver can be written in the form

$$V=\frac{B}{w}UG(ks). \tag{7}$$

Similarly, the acoustic pressure radiated by the MHD transmitter is equal to

$$p_{\text{out}}=\frac{B}{w}IG(ks). \tag{8}$$

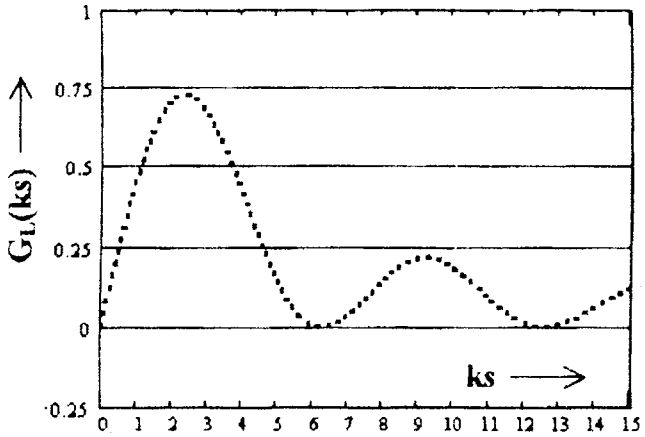


FIG. 1. Transfer function $G_L(ks)$ of MHD transmitter for rigid backing condition.

a)Electronic mail: ado@zakus.ita.pwr.wroc.pl

Then, the transduction coefficient T is given by the following equation:

$$T = \frac{B}{w} G(ks). \tag{9}$$

The transfer function $G(ks)$ is complex in the general case. It depends on the boundary condition. For a thin layer of electrolyte, the transduction coefficient is equal to B/w [see Eq. (15) and the text before Eq. (34) in Ref. 1]. The frequency dependence of total transduction coefficient, described by the transfer functions $G(ks)$, is caused by interference of waves between $z=0$ and $z=s$.

[1] A. J. Campanella, "Investigations of sound waves generated by the Hall effect in electrolytes," J. Acoust. Soc. Am. **111**, 2087–2096 (2002).
[2] S. Ballantine, "Reciprocity in electromagnetic, mechanical, acoustical, and interconnected systems," Inst. Radio Eng. Proc. **17**, 929–951 (1929).

Reply to "Comment on 'Investigation of sound waves generated by the Hall effect in electrolytes'" [J. Acoust. Soc. Am. 113(3), 1207–1208 (2003)] (L)

Angelo J. Campanella
3201 Ridgewood Drive, Columbus (Hilliard), Ohio 43026

(Received 17 November 2002; accepted for publication 16 December 2002)

[DOI: 10.1121/1.1543549]

PACS numbers: 43.38.Ar, 43.38.Dv, 43.35.Rw [AJZ]

The "rigid" curve in Fig. 5(a) was printed in error. Equation (27) indicates the form of the rigid backing response, $G_L(ks)$, as suggested by Dobrucki. The revised figure, attached, should replace the originally published Fig. 5(a). The "rigid" curve in Figs. 7(a) and (b) is correct.

The factor of two appearing in (38) and (39) is verified as follows. Introduce C in Eq. (35) as the amplitude for the $-z$-going wave. After integration (34), Eq. (36) becomes

$$V = Ke^{j\omega t}[C(e^{jks}-1) - D(e^{-jks}-1)], \tag{36}$$

where $K = hB\xi_{,to}/(jks)$. For a matched impedance condition, such as when the electrodes are immersed in a continuous conducting medium, there is no returning wave, so that $D=0$. The amplitude, C, will be unity when K contains the physical quantities appropriate to reciprocal MHD action. Then

$$\begin{aligned} V &= Ke^{j\omega t}[e^{jks}-1] \\ &= Ke^{j\omega t}e^{-jks/2}[e^{jks/2}-e^{-jks/2}] \\ &= 2Ke^{j\omega t}e^{-jks/2}[\sin(ks/2)], \end{aligned}$$

$$V = hB\xi_{,to}e^{-jks/2}[\sin(ks/2)]/(jks/2) \quad \text{(matched)}.$$

For a rigid boundary at $z=0$, then $u_{z=0}=0$, and $D=-C$, so that the reflected wave amplitude can be equal in amplitude but opposite in phase to the incident wave. Equation (36), applied to the rigid backing case, provides

$$\begin{aligned} V &= Ke^{j\omega t}[(e^{jks}-1)+(e^{-jks}-1)] \\ &= 2Ke^{j\omega t}[1/2(e^{jks}+e^{-jks})-1)] \\ &= 2Ke^{j\omega t}[\cos(ks)-1] \end{aligned} \tag{36a}$$

$$V = 2jhB\xi_{,to}e^{j\omega t}[1-\cos(ks)]/(ks) \tag{38}$$

[rigid, as shown in Fig. 5(b)].

The desire for parity in reciprocity amplitudes might be resolved by a closer examination of the amplitude assumptions associated with (14) and consideration that the MHP process is inefficient.

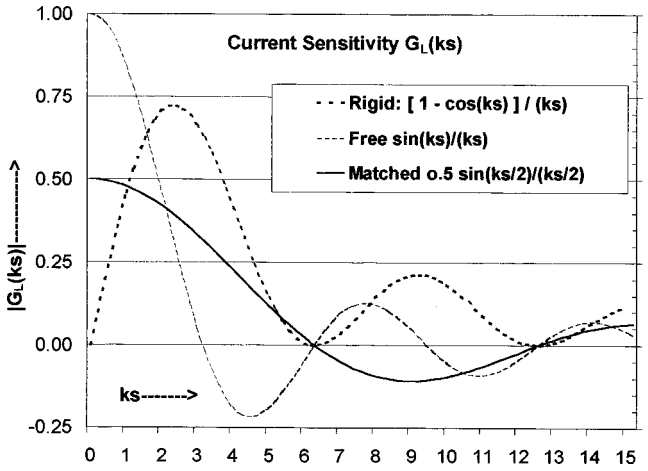


FIG. 5(a).

 0001-4966/2003/113(3)/1209/1/$19.00

Impedance measurements around grazing incidence for nonlocally reacting thin porous layers

Jean-François Allard,[a] Michel Henry, and Vincent Gareton
Laboratoire d'Acoustique de l'Université du Maine, UMR CNRS 6613, Avenue Olivier Messiaen, 72085 Le Mans Cedex 9, France

Gert Jansens and Walter Lauriks
Laboratorium voor Akoestiek en Thermische Fysica, Departement Natuurkunde, Katholieke Universiteit Leuven, Celestijnenlaan 200 D, B-3001 Heverlee, Belgium

(Received 24 May 2001; revised 6 August 2002; accepted 13 November 2002)

For locally reacting materials with a constant surface impedance, a classical method based on the work of Chien and Soroka [J. Sound Vib. **43**, 9–20 (1975)] for measuring this impedance *in situ* around grazing incidence is currently used. A generalization of this work to include thin nonlocally reacting materials with a surface impedance noticeably dependent on the angle of incidence is performed. It is shown that the model by Chien and Soroka can be used, though the constant surface impedance must be replaced by the impedance at grazing incidence for the evaluation of the numerical distance. Measurements performed on a thin porous layer using this method are compared with measurements performed using the near-field acoustical holography method [M. Tamura, J. Acoust. Soc. Am. **88**, 2259–2264 (1990)]. Other measurements performed on a fibrous layer are in good agreement with the predicted values of the impedance at grazing incidence. © *2003 American Institute of Physics.* [DOI: 10.1121/1.1536610]

PACS numbers: 43.20.El, 43.58.Bh, 43.28.En [ANN]

I. INTRODUCTION

Recent work has been performed concerning the measurement of the surface impedance of porous surfaces at normal incidence in an impedance tube[1] and at oblique incidence in a free field.[2–5] At oblique incidence, the evaluation of the surface impedance and reflection coefficient can be performed from measurements of the monopole pressure field with an expression of the reflected monopole field p_r obtained by Chien and Soroka[6] under the condition that the surface impedance Z_s does not depend on the angle of incidence. The source and the receiver are represented in Fig. 1. Let R_1 and R_2 be the distance between the receiver at M and the image of the source S', and the receiver and the source S, respectively. Let θ_0 be the specular angle of incidence, ν the reflection coefficient, and k the wave number in air. In Ref. 6, the reflected field p_r is written for $kR_1 \gg 1$ [see Eq. (25) of Ref. 6, with the time dependence $\exp(-i\omega t)$],

$$p_r = \frac{\exp(ikR_1)}{R_1}[\nu(\sin\ \theta_0) + (1-\nu(\sin\ \theta_0))F(p_e)], \quad (1)$$

where $F(p_e)$ is given by

$$F(p_e) = 1 + i\sqrt{\pi}p_e\ \exp(-p_e^2)\mathrm{erfc}(-ip_e). \quad (2)$$

The numerical distance p_e, defined in Ref. 3, is given by

$$p_e = \left(\frac{1}{2}ikR_1\right)^{1/2}\left(\frac{Z}{Z_s} + \cos\ \theta_0\right), \quad (3)$$

where Z is the characteristic impedance of air, and the plane wave reflection coefficient ν is given by

$$\nu(\sin\ \theta_0) = (\cos\ \theta_0 - Z/Z_s)/(\cos\ \theta_0 + Z/Z_s). \quad (4)$$

Poles of the reflection coefficient exist for an incidence angle θ_p satisfying

$$\cos\ \theta_p = -Z/Z_s. \quad (5)$$

Equation (1) is valid under two conditions, $kR_1 \gg 1$, and θ_p close to the stationary point θ_0 on the integration path. The latter condition is satisfied around grazing incidence if $|Z_s| \gg Z$. Equation (1) has been extensively used in the context of outdoors sound propagation over grounds having a large surface impedance Z_s. A review of the works has been performed by Attenborough.[7] For a porous layer set on a rigid impervious surface, Z_s, which is defined for an incident plane wave, depends on the angle of incidence θ. If the porous frame is motionless, Z_s is given by [see Eq. (5.98) of Ref. 8 with the time dependence $\exp(i\omega t)$]

$$Z_s(\theta) = i\frac{Z_1}{\phi\ \cos\ \theta_1}\cot(k_1 l\ \cos\ \theta_1). \quad (6)$$

In Eq. (6), θ_1 is the complex angle of refraction, ϕ is the porosity, k_1 is the wave number, Z_1 is the characteristic impedance in the air saturating the medium, and l is the thickness of the layer. A thin porous layer is characterized by $|k_1|l \ll 1$, and the porous frame is generally motionless with a good approximation when this condition is fulfilled.[9] The cosine of the refraction angle is given by [see Eq. (5.98) of Ref. 8]

$$\cos\ \theta_1 = \left(1 - \frac{1}{n^2} + \frac{\cos^2\ \theta}{n^2}\right)^{1/2}, \quad (7)$$

where n is the refraction index, $n = k_1/k$. For the usual sound absorbing media such as low flow-resistivity foams and glass

[a]Electronic mail: Jean-francois.allard@univ-lemans.fr

 0001-4966/2003/113(3)/1210/6/$19.00

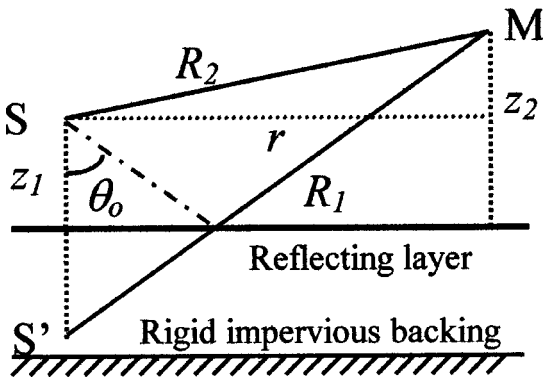


FIG. 1. The source S, its image S' and the receiver at M over the porous layer, $SM=R_2$, $S'M=R_1$, r is the horizontal distance between S and M, and θ_0 the specular reflection angle.

wools, $|n|$ above 1 kHz lies between 1 and 1.5, $\cos\theta$ and Z_s can noticeably depend on the angle of incidence, and the model by Chien and Soroka[6] is not valid. Li *et al.*[2] have developed a heuristic model for the case of nonlocally reacting media. The one modification of the model by Chien and Soroka[6] is the substitution of $Z_s(\theta_0)$ for the constant impedance Z_s in Eq. (3).

In the Appendix, the general formulation of the "reference integral method" by Brekhovskikh and Godin[10] is used to adapt the pole subtraction method for the case of nonlocally reacting media. It is shown in Sec. II that for θ_0 close to $\pi/2$ and $|Z_s(\pi/2)|\gg Z$, the general expression obtained in the appendix can be simplified and $Z_s(\pi/2)$ can be substituted for Z_s in Eq. (3). A simple experimental procedure to evaluate $Z_s(\pi/2)$ from pressure measurements performed at small source–receiver distances is suggested. Measurements are presented in Sec. III.

II. THE REFLECTED MONOPOLE FIELD

Let z_1 and z_2 be the distance between the porous surface and the source S, and the receiver at M, respectively (see Fig. 1). The plane wave reflection coefficient is given by

$$\nu(\sin\theta)=(Z_s(\theta)-Z/\cos\theta)/(Z_s(\theta)+Z/\cos\theta), \tag{8}$$

which can be rewritten

$$\nu(q)=\frac{-\sqrt{n^2-q^2}+i\dfrac{Z_1}{\phi Z}n\sqrt{1-q^2}\cot(kl\sqrt{n^2-q^2})}{\sqrt{n^2-q^2}+i\dfrac{Z_1}{\phi Z}n\sqrt{1-q^2}\cot(kl\sqrt{n^2-q^2})}, \tag{9}$$

where $q=\sin\theta$. Poles of the reflection coefficient are solutions of the implicit equation

$$\cos\theta_p=-Z/Z_s(\theta_p). \tag{10}$$

Using the Sommerfeld representation, the reflected field is given by[10]

$$p_r=ik\int_0^\infty \frac{q\,dq}{\mu}\,\nu(q)J_0(kqr)\exp[ik\mu(z_1+z_2)], \tag{11}$$

where $\mu=\sqrt{1-q^2}$, $\mathrm{Im}\,\mu\geqslant 0$, and $r=R_1\sin\theta_0$ (see Fig. 1).

The integration is carried out on the real $q=\sin\theta$ axis in the complex $\sin\theta$ plane. For $kR_2\gg 1$, the steepest descent method can be used to evaluate the integral. The stationary point on the passage path is located at $\sin\theta_0$. It can be demonstrated[11] that for thin porous layers a pole of the reflection coefficient exists at $\sin\theta_p$ close to 1. A more precise localization of the pole can be obtained in the following way. For θ close to $\pi/2$, $|\cos\theta|\ll 1$, $\cos\theta_1$ given by Eq. (7) is stationary at $(1-1/n^2)^{1/2}$, and $Z_s(\theta)$ can be replaced by $Z_s(\pi/2)$. From Eq. (10), a pole of the reflection coefficient at θ_p close to $\pi/2$ must satisfy with a good approximation

$$\cos\theta_p=-\frac{Z}{Z_s(\pi/2)}. \tag{12}$$

Equation (12) actually gives a solution θ_p close to $\pi/2$. For a thin porous layer characterized by $|k_1|l\ll 1$, $Z/Z_s(\pi/2)\ll 1$.

An iterative process can be used to obtain the exact value of $\cos\theta_p$. The stationary point located at $\sin\theta_0$ is close to $\sin\theta_p$ for angles of specular reflection close to $\pi/2$. The classical passage path method must be modified and replaced by the pole subtraction method. The main steps of the evaluation of p_r are given in the Appendix. This procedure leads to the simplified expression which is valid at θ_0 close to $\pi/2$ and for thin porous layers [see Eq. (A16), $\sin\theta_0$ and $\sin\theta_p$ being close to 1 and the factor $(\sin(\theta_0)/\sin(\theta_p))^{1/2}$ being removed]

$$p_r=\frac{\exp(ikR_1)}{R_1}\{\nu(\sin\theta_0)+(1-\nu(\sin\theta_0))\times[1+i\sqrt{\pi}p'_e\exp(-p'^2_e)\mathrm{erfc}(-ip'_e)]\}, \tag{13}$$

where p'_e is given by

$$p'_e=\left(\frac{1}{2}ikR_1\right)^{1/2}\left(\frac{Z}{Z_s(\pi/2)}+\cos\theta_0\right). \tag{14}$$

Equations (13) and (14) are similar to Eqs. (1)–(3), but the constant impedance Z_s in Eq. (3) is now replaced by $Z_s(\pi/2)$ in Eq. (14). A similar result is obtained with the model by Li *et al.*[2] for θ_0 close to $\pi/2$. In the context of this model, Z_s in Eq. (3) is replaced by $Z_s(\theta_0)$, which is close to $Z_s(\pi/2)$ as indicated previously. A good agreement is obtained in Ref. 2 at grazing incidence for a thin porous layer between the attenuation predicted with the model of Li *et al.*[2] and with the fast field formulation (see Fig. 3-b of Ref. 2). Equation (13) can be rewritten as shown in the Appendix [see Eq. (A19)]

$$p_r=\frac{\exp(ikR_1)}{R_1}\left[1-\sqrt{2\pi kR_1}\exp\left(\frac{3\pi i}{4}\right)\times\cos\theta_p\exp(-p'^2_e)\mathrm{erfc}(-ip'_e)\right]. \tag{15}$$

For *in situ* measurements on sound absorbing materials, the distance from the source to the receiver must be as small as possible to diminish the relative contribution of the diffuse field, the modulus of the numerical distance will generally be smaller than 1, and $\mathrm{erfc}(-ip'_e)$ can be replaced by the serial development

$$\mathrm{erfc}(-ip'_e)=1+\frac{2i}{\sqrt{\pi}}p'_e\left(1+\frac{p'^2_e}{3}+\frac{p'^4_e}{2!5}+\frac{p'^6_e}{3!7}+\ldots\right). \tag{16}$$

A first approximation for the evaluation of $\cos\theta_p$ is obtained from the 0 order approximation in p'_e of $\mathrm{erfc}(-ip'_e)\times\exp(-p'^2_e)$,

$$\cos\theta_p=\frac{p_h-p_r}{p_h}\exp\left(-\frac{3\pi i}{4}\right)\Big/(2\pi kR_1)^{1/2}. \tag{17}$$

In Eq. (17), $p_h=\exp(ikR_1)/R_1$ is the field reflected by a hard surface and p_r is the actual reflected field. An iterative procedure can be used from this first approximation to obtain the exact solution of Eq. (15) that satisfies

$$\cos\theta_p=\frac{p_h-p_r}{p_h}\exp\left(-\frac{3\pi i}{4}\right)\Big/[(2\pi kR_1)^{1/2}\times\exp(-p'^2_e)\mathrm{erfc}(-ip'_e)], \tag{18}$$

$$p'_e=(\tfrac{1}{2}ikR_1)^{1/2}(-\cos\theta_p+\cos\theta_0). \tag{19}$$

A simple evaluation of $Z_s(\pi/2)$ is obtained from Eq. (12). Previous methods generally require the measurement of the direct field and the field over the porous surface. The impedance deduction technique is generally not very straightforward. With the present method, a first approximation is obtained for $\cos\theta_p$ for each source–receiver geometry from two measurements, successively over a hard impervious surface and over the porous surface. The use of the first approximation obtained with Eq. (17) as an initial value for $\cos\theta_p$ in the iterative process leads to a simple procedure for the evaluation of $\cos\theta_p$. For a given source–receiver geometry, the use of the method is limited to a range of frequencies where kR_1 simultaneously is sufficiently large for the asymptotic estimation of p_r from Eq. (13) to be valid and sufficiently small for the first approximation of $\cos\theta_p$ by Eq. (17) to be a good initial value for the iterative process. Simulations performed with p_r evaluated from Eq. (11) show that this procedure can be used for $kR_1>4$, allowing impedance measurements for source-receiver distances larger than 20 cm around 1 kHz.

III. MEASUREMENTS

Measurements have been performed for two materials, at 1.5 and 2 kHz for a reticulated foam of thickness $l=0.5$ cm, with lateral dimensions 1 m × 1 m, and at 0.75 and 1 kHz for a fibrous layer of thickness $l=1.25$ cm, with lateral dimensions 2 m × 2 m. The fibrous material has been previously studied by Brouard *et al.*,[12] and the macroscopic parameters that characterize sound propagation in the material are given in Table 1 of Ref. 12. Measurements around grazing incidence have been performed in an anechoic room for source–microphone distances $R_1=0.2$, 0.3, 0.4, 0.5, and 0.6 m. The distances z_1 and z_2 are equal, $z_1=z_2=4$ mm. The source is a driver unit feeding a pipe having an internal diameter $d=6$ mm. Measurements of the transfer functions between the input signal feeding the driver unit and the output signal of the microphone have been performed above the porous layer and above a rigid impervious surface. Let T_h and T_r be the transfer functions for the rigid impervious surface and the porous layer, respectively. If the amplitude of the source depends only on the input signal and not on the nature of the reflecting surface, $(p_h-p_r)/p_h$ is given with a good approximation by

$$\frac{p_h-p_r}{p_h}=\frac{2(T_h-T_r)}{T_h}. \tag{20}$$

For each source-receiver geometry, a first approximation for $\cos\theta_p$ is obtained from Eq. (17), the approximation being used as an initial value to obtain the solution of Eqs. (18) and (19) by an iterative process. For the foam at 1.5 kHz, a small but systematic variation of the evaluated $\mathrm{Re}(\cos\theta_p)$ as a function of R_1 appears, for $R_1=0.2$ m, $\cos\theta_p=-0.094+i0.13$ and for $R_1=0.6$ m, $\cos\theta_p=-0.080+i0.13$. This could be due to a modification of the source signal at constant input signal when the porous surface is replaced by the rigid impervious surface, the source being close to the reflecting surface. A complex parameter α close to 1 has been used to take into account the modification of the amplitude

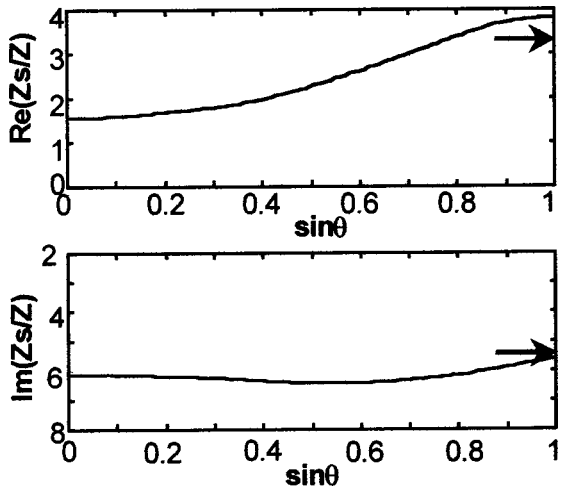


FIG. 2. The surface impedance of the foam at 1500 Hz, measured with the NAH method from normal to grazing incidence. The arrows indicate the evaluation using the new method.

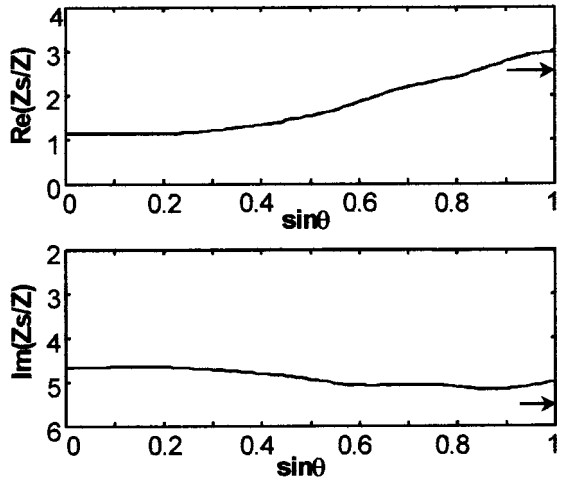


FIG. 3. The surface impedance of the foam at 2000 Hz, measured with the NAH method from normal to grazing incidence. The arrows indicate the evaluation using the new method.

TABLE I. Normalized surface impedance of the foam at grazing incidence obtained using the NAH method and the new method.

	NAH	New method
1.5 kHz	$3.75\pm0.5+i$ (5.6 ± 0.8)	$3.3\pm0.3+i$ (5.5 ± 0.3)
2 kHz	$2.95\pm0.5+i$ (5.0 ± 0.6)	$2.6\pm0.2+i$ (5.6 ± 0.2)

of the source when the rigid surface is replaced by the porous layer. This parameter is adjusted at 1500 Hz by removing the systematic variations on the evaluation of $\cos\theta_p$, $(p_h - p_r)/p_h$ in Eqs. (17) and (18) being replaced by $2(\alpha T_h - T_r)\alpha T_h$. With $\alpha=1-0.1i$, for $R_1=0.2$ m, $\cos\theta_p = -0.081+i0.14$ and for $R_1=0.6$ m, $\cos\theta_p=-0.079 + i0.13$.

For the foam, measurements have also been performed with the near-field acoustical holography (NAH) method[13,14] from 0 to $\pi/2$. Measurements with the NAH method are presented in Fig. 2 at 1500 Hz and Fig. 3 at 2000 Hz. The arrows at $\theta=\pi/2$ correspond to the mean value of the estimations for the six locations of the receiver with the new method and $\alpha=1-0.1i$. For the sake of simplicity, the same coefficient α has been used for all evaluations in what follows.

In Table I, evaluations at 1500 and 2000 Hz of $Z_s(\pi/2)/Z$ with both methods are indicated, with errors estimated from the dispersion of the different measurements. Measurements with the NAH method at $\theta=\pi/2$ and with the new method are compatible. Nevertheless, measurements with the NAH method do not give precise measurements on poor sound-absorbing materials at grazing incidence. A better estimation of the performances of the new method should be obtained by comparing measurements with predictions obtained on a medium where a precise determination of n and Z_1 has been performed. For the fibrous material previously studied in Ref. 12, the surface impedance at $\theta=\pi/2$ has been measured with the new method at 0.75 and 1 kHz, and comparisons have been performed with predictions obtained from Eq. (9) for $q=1$ and $q=0$. The model and the macroscopic parameters used to predict n and Z_1 are given in Ref. 12. Predictions at $\theta=\pi/2$ and measurements with the new method are given in Table II. For the measurements with the new method, the same source–receiver geometry and the same $\alpha=1-0.1i$ as for the foam have been used. Measurements and predictions are compatible. At normal incidence, the predicted reduced impedances are $0.73+i4.63$ at 0.75 kHz and $0.64+i3.5$ at 1 kHz, which is far from the measured values of Table II.

IV. CONCLUSION

The model implemented in the present work predicts that the classical expression by Chien and Soroka remains valid for thin nonlocally reacting materials, though the constant impedance related to locally reacting media must be replaced by the impedance at grazing incidence for the evaluation of the numerical distance. This interpolation is compatible with measurements performed using the NAH method. Measurements at grazing incidence can be performed for small source–receiver distances, at least in the medium and the high frequency ranges. *In situ* measurements for poor sound-absorbing materials should be possible with this method.

TABLE II. Normalized surface impedance of the fibrous material at grazing incidence predicted and measured using the new method.

	Prediction $\theta=\pi/2$	New method
0.75 kHz	$1.38+i$ 4.62	$1.26\pm0.2+i(4.3\pm0.4)$
1 kHz	$1.30+i$ 3.50	$1.21\pm0.2+i(3.14\pm0.3)$

APPENDIX

The reflected pressure field given by Eq. (11) can be rewritten, taking into account the relation between the 0 order Bessel function and the Hankel function of the first kind $J_0(u)=0.5(H_0^{(1)}(u)-H_0^{(1)}(-u))$ [see Eq. (1.1.10) of Ref. 10]

$$p_r=\frac{ik}{2}\int_{-\infty}^{+\infty}\frac{q\,dq}{\mu}\,\nu(q)H_0^{(1)}(kr)\exp(ik\mu(z_1+z_2)), \quad \text{(A1)}$$

with the following expression for v

$$\nu(q)=1-2\sqrt{n^2-q^2}\Big/\left(\sqrt{n^2-q^2}+i\frac{Z_1}{\phi Z}n\sqrt{1-q^2}\cot k(l\sqrt{n^2-q^2})\right) \quad \text{(A2)}$$

obtained from Eq. (9).

The unit term in Eq. (A2) provides in Eq. (A1) a contribution $p_1=\exp(ikR_1)/R_1$ to the reflected field. The contribution p_2 of $\nu(\sin\theta)-1$ is calculated following Brekhovskikh and Godin[5] by using the asymptotic expansion of the Hankel function and can be written [see Eqs. (1.2.2)–(1.2.3) of Ref. 10]

$$p_2=\left(\frac{k}{2\pi r}\right)^{1/2}\exp\left(\frac{i\pi}{4}\right)\int_{-\infty}^{+\infty}F(q)\exp[kR_1f(q)]\,dq, \quad \text{(A3)}$$

$$f(q)=i[q\sin\theta_0+(1-q^2)^{1/2}\cos\theta_0], \quad \text{(A4)}$$

$$F(q)=-2\sqrt{\frac{q}{1-q^2}}\sqrt{n^2-q^2}\left(\sqrt{n^2-q^2}+i\frac{Z_1}{\phi Z}n\sqrt{1-q^2}\cot k(l\sqrt{n^2-q^2})\right)^{-1}. \quad \text{(A5)}$$

The integral is performed in the complex $\sin\theta$ plane on the real axis. Following Brekhovskikh and Godin, the integration will be performed on the steepest descent path, which is defined by $q\sin\theta_0+(1-q^2)^{1/2}\cos\theta_0=1+is^2$, $-\infty<s<\infty$ [see Eq. (1.2.3) of Ref. 10].

The pole of v and the stationary point on the path are located in the $\sin\theta$ plane at $q_p=\sin\theta_p$ and $q_0=\sin\theta_0$. The asymptotic expression I for the integral in Eq. (A3), valid for kR_1 sufficiently large and q_p sufficiently close to each other, is [see Eqs. (A.3.12)–(A.3.13) of Ref. 10]

$$I=\exp(kR_1 f(q_0))\left[aF_1(1,kR_1,s_p)+\left(\frac{\pi}{kR_1}\right)^{1/2}\phi_1(0)\right]. \tag{A6}$$

In Eq. (6), ϕ_1, F_1, a, and s_p are given by

$$\phi_1(0)=F(q_0)\sqrt{-2/f''(q_0)}+a/s_p, \tag{A7}$$

$$F_1(1,kR_1,s_p)=i\pi\exp(-kR_1 s_p^2)\operatorname{erfc}(-i\sqrt{kR_1}s_p), \tag{A8}$$

$$a=\lim_{q\to q_p} F(q)(q-q_p), \tag{A9}$$

$$s_p=\exp\left(\frac{i\pi}{4}\right)\sqrt{2}\sin\frac{\theta_p-\theta_0}{2}. \tag{A10}$$

Let $G(q)$ be the factor in braces in Eq. (A5). For a pole of the first order, the parameter a is given by

$$a=-2\sqrt{\frac{q_p}{1-q_p^2}}\frac{\sqrt{n^2-q^2}}{G_q'(q_p)}, \tag{A11}$$

where $G_q'(q_p)$ is the sum of three terms

$$G_p'(q_p)=-\frac{k}{k_1}\frac{\sin\theta_p}{\cos\theta_{1p}}+\frac{inZ_1}{\phi Z}\left[-\tan\theta_p\cot(k_1 l\cos\theta_{1p})+\frac{\cos\theta_p}{\sin^2(k_1 l\cos\theta_{1p})}\frac{k^2 l\sin\theta_p}{k_1\cos\theta_{1p}}\right]. \tag{A12}$$

When $|k_1 l|$ tends to 0, the second term on the right hand side of Eq. (A12) tends to ∞, and the first and the third terms have a finite limit. The first and the third terms are discarded under the hypothesis $|k_1 l|\ll 1$, and a is given by

$$a=2Z/(\sqrt{\sin\theta_p}Z_s(\theta_p)). \tag{A13}$$

In Eq. (A7), $f''(q_0)=-i/\cos^2\theta_0$, and in Eq. (A6), $\phi_1(0)$ is given by

$$\phi_1(0)=[-1+\nu(\cos\theta_0)]\sqrt{\frac{q_0}{1-q_0^2}}\sqrt{-2i\cos^2\theta_0}+\frac{2Z}{Z_s(\theta_p)\sqrt{\sin\theta_p}}\frac{1}{s_p}, \tag{A14}$$

and p_r can be written

$$p_r=\frac{\exp(ikR_1)}{R_1}+\left(\frac{k}{2\pi r}\right)^{1/2}\exp\left(\frac{i\pi}{4}\right)\exp(ikR_1)\{A+B\},$$

with

$$A=\frac{2Z}{Z_s(\theta_p)\sqrt{\sin\theta_p}}[i\pi\exp(-kR_1 s_p^2)\operatorname{erfc}(-i\sqrt{kR_1}s_p)], \tag{A15}$$

and

$$B=\left(\frac{\pi}{kR_1}\right)^{1/2}\left[\sqrt{\frac{q_0}{1-q_0^2}}(-1+\nu(\sin\theta_0))\sqrt{-2i\cos^2\theta_0}+\frac{2Z}{Z_s(\theta_p)\sqrt{\sin\theta_p}s_p}\right].$$

After some rearrangement, this equation becomes

$$p_r=\frac{\exp(ikR_1)}{R_1}\times\left\{v(\sin\theta_0)+\frac{Z\sqrt{\sin\theta_0}}{Z_s(\theta_p)\sqrt{\sin\theta_p}\sin\frac{\theta_p-\theta_0}{2}}C\right\}, \tag{A16}$$

with

$$C=[1+i\sqrt{\pi}s_p\sqrt{kR_1}\exp(-kR_1 s_p^2)\operatorname{erfc}(-i\sqrt{kR_1}s_p)].$$

For thin porous layers, $\sqrt{\sin\theta_p}$ is very close to 1, and for θ_0 close to $\pi/2$, $\sqrt{\sin\theta_0}$ is also very close to 1, and the two terms can be removed. Moreover, if θ_p and θ_0 are close to $\pi/2$, $\sin(\theta_p-\theta_0)/2$ at the second order approximation in $\theta_p-\pi/2$ and $\theta_0-\pi/2$ can be replaced by $(\cos\theta_0-\cos\theta_p)/2$. The new numerical distance $p_e'=\sqrt{kR_1}s_p$ can be written

$$p_e'=\left(\frac{1}{2}ikR_1\right)^{1/2}\left(\frac{Z}{Z_s(\pi/2)}+\cos\theta_0\right). \tag{A17}$$

In Eq. (A16), $Z/(Z_s(\theta_p)\sin(\theta_p-\theta_0)/2)$ can be replaced by $1-v(\sin\theta_0)$ and the simplified equation (A16) becomes

$$p_r=\frac{\exp(ikR_1)}{R_1}\{\nu(\sin\theta_0)+(1-\nu(\sin\theta_0))\times[1+ip_e'\exp(-p_e'^2)\operatorname{erfc}(-ip_e')]\}, \tag{A18}$$

which can be rewritten

$$p_r=\frac{\exp(ikR_1)}{R_1}\left\{1-\sqrt{2\pi kR_1}\exp\left(\frac{3\pi i}{4}\right)\cos\theta_p\times\exp(-p_e'^2)\operatorname{erfc}(-ip_e')\right\}. \tag{A19}$$

[1] C. W. Beamer and R. T. Muehleisen, "Comparison of techniques for measuring the acoustic properties of porous materials," J. Acoust. Soc. Am. **108**, 2519 (2000).

[2] K. M. Li, T. Waters-Fuller, and K. Attenborough, "Sound propagation from a point source over extended-reaction ground," J. Acoust. Soc. Am. **104**, 679–685 (1998).

[3] C. Nocke, H. V. Fuchs, and V. Mellert, "Acoustic impedance determination of large absorbent linings," Euro-noise 98, München (1998), pp. 853–858.

[4] C. Nocke, "Improved impedance deduction from measurement near grazing incidence," Acust. Acta Acust. **85**, 586–590 (1999).

[5] S. Taherzadeh and K. Attenborough, "Deduction of ground impedance from measurements of excess attenuation spectra," J. Acoust. Soc. Am. **105**, 2039–2042 (1999).

[6] C. F. Chien and W. W. Soroka, "Sound propagation along an impedance plane," J. Sound Vib. **43**, 9–20 (1975).

[7] K. Attenborough, "Review of ground effects on outdoor sound propagation from continuous broadband sources," Appl. Acoust. **24**, 289–319 (1988).

[8] J. F. Allard, *Propagation of Sound in Porous Media: Modeling Sound*

Absorbing Materials (Chapman & Hall, London, 1993).
[9] A. Bardot, B. Brouard, and J. F. Allard, ''Frame decoupling at low frequencies in thin porous layers saturated by air,'' J. Appl. Phys. **79**, 8223–8229 (1996).
[10] L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media II, Point Source and Bounded Beams*, Springer Series on Wave Phenomena (Springer, New York, 1992).
[11] W. Lauriks, L. Kelders, and J. F. Allard, ''Poles and zeros of the reflection coefficient of a porous layer having a motionless frame in contact with air,'' Wave Motion **28**, 59–67 (1998).
[12] B. Brouard, D. Lafarge, and J. F. Allard, ''Measurement and prediction of the surface impedance of a resonant sound absorbing structure,'' Acta Acustica **2**, 301–306 (1994).
[13] M. Tamura, ''Spatial Fourier transform method of measuring reflection coefficient at oblique incidence,'' J. Acoust. Soc. Am. **88**, 2259–2264 (1990).
[14] B. Brouard, D. Lafarge, J. F. Allard, and M. Tamura, ''Measurement and prediction of the reflection coefficient of porous layers at oblique incidence and for inhomogeneous waves,'' J. Acoust. Soc. Am. **99**, 100–107 (1996).

Near-field calculations for a rigid spheroid with an arbitrary incident acoustic field

John P. Barton,[a] Nicholas L. Wolff, Haifeng Zhang, and Constantine Tarawneh
Department of Mechanical Engineering, College of Engineering & Technology, University of Nebraska—Lincoln, Lincoln, Nebraska 68588-0656

(Received 15 February 2002; revised 8 November 2002; accepted 25 November 2002)

A general spheroidal coordinate separation-of-variables solution is developed for the determination of the acoustic pressure distribution near the surface of a rigid spheroid for a monofrequency incident acoustic field of arbitrary character. Calculations are presented, for both the prolate and oblate geometries, demonstrating the effects of incident field orientation and character (plane-wave, spherical wave, cylindrical wave, and focused beam) on the resultant acoustic pressure distribution.
 [DOI: 10.1121/1.1538200]

PACS numbers: 43.20.El, 43.20.Fn, 43.20.Bi, 43.35.Ae [LLT]

I. INTRODUCTION

The theoretical determination of the acoustic scattering that results from the interaction of an incident plane wave or point source with a rigid, fixed-position prolate spheroid has been previously considered. The problem has been addressed using the spheroidal coordinate separation-of-variables approach,[1–9] using the spherical coordinate extended boundary method (sometimes equivalently referred to as the T-matrix method),[10–12] and, recently, using a cylindrical function deformed cylinder method.[13] The rigid oblate spheroid scattering problem has apparently drawn less interest.[14] These earlier works were predominantly concerned with farfield scattering or, at most, the determination of the acoustic pressure at the surface of the spheroid. In addition, many of these earlier works are not general solutions, but are restricted in some way, requiring end-on incidence or a high-frequency/high aspect ratio asymptotic limit.

The study presented here, a spheroidal coordinate separation-of-variables solution of the rigid spheroid acoustic scattering problem, differs from earlier efforts in that (1) the incident field can be any known monofrequency field (e.g., a focused acoustic beam or the scattered field from an adjacent object), not just a plane wave or a point source and (2) the solution for the oblate spheroid, as well as the prolate spheroid, is included.

II. THEORETICAL DEVELOPMENT

The interaction of a known incident monofrequency acoustic field with a motionless, rigid spheroid is considered. The medium surrounding the spheroid is assumed to be homogeneous, infinite, and nonabsorbing. A schematic of the geometrical arrangement is given in Fig. 1. An implicit time dependence of $\exp(-i\omega t)$ is assumed and the acoustic pressure is nondimensionalized relative to a reference acoustic pressure amplitude associated with the incident field (p_r).

A spheroidal surface can be created by rotating an ellipse about its axis. If the ellipse is rotated about its major axis, a prolate spheroid is formed. If the ellipse is rotated about its minor axis, an oblate spheroid is formed. If a is the length of the semimajor axis and b is the length of the semiminor axis, then

$$f=a[1-(b/a)^2]^{1/2}, \tag{1}$$

where f is the semifocal length of the spheroid. In the following, the rectangular coordinates (x,y,z), the spatial position vector $\mathbf{r}$, and the vector calculus operators (∇,∇^2) are all nondimensionalized relative to the semifocal length f.

In prolate coordinates, the normalized rectangular coordinates are related to the spheroidal coordinates $(-1<\eta<+1, 1<\xi<\infty, 0<\phi<2\pi)$ by

$$x=[(\xi^2-1)(1-\eta^2)]^{1/2}\cos\phi, \tag{2}$$

$$y=[(\xi^2-1)(1-\eta^2)]^{1/2}\sin\phi, \tag{3}$$

$$z=\xi\eta, \tag{4}$$

where ξ is the spheroidal radial coordinate, η is the spheroidal angle coordinate, and ϕ is the azimuthal angle. A surface of constant $\xi=\xi_0$ forms the surface of a prolate spheroid with an axis ratio of

$$(a/b)=\frac{\xi_0}{(\xi_0^2-1)^{1/2}}. \tag{5}$$

So, for example, if $\xi_0=1.15470$ then $(a/b)=2$, if $\xi_0=1.060\,66$ then $(a/b)=3$, etc.

In oblate coordinates, the normalized rectangular coordinates are related to the spheroidal coordinates $(-1<\eta<+1, 0<\xi<\infty, 0<\phi<2\pi)$ by

$$x=[(\xi^2+1)(1-\eta^2)]^{1/2}\cos\phi, \tag{6}$$

$$y=[(\xi^2+1)(1-\eta^2)]^{1/2}\sin\phi, \tag{7}$$

$$z=\xi\eta. \tag{8}$$

In this case, a surface of constant $\xi=\xi_0$ forms the surface of an oblate spheroid with an axis ratio of

$$(a/b)=[1+(1/\xi_0^2)]^{1/2}. \tag{9}$$

So, for example, if $\xi_0=0.577\,35$ then $(a/b)=2$, if $\xi_0=0.353\,55$, then $(a/b)=3$, etc.

[a]Electronic mail: jbarton@unlserve.unl.edu

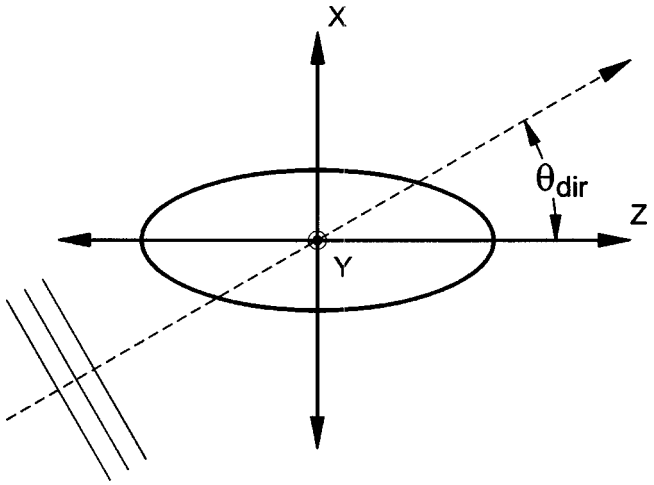


FIG. 1. Schematic of the geometrical arrangement. The spheroid is axisymmetric about the z axis. For the prolate spheroid (as shown) the major axis is along the z axis. For an oblate spheroid, the major axis would be perpendicular to the z axis.

The eigenfunctions of the spheroidal coordinate separation-of-variables solution of the time-independent wave equation (Helmholtz equation),

$$\nabla^2 p + h^2 p = 0, \tag{10}$$

are of the form

$$p_{lm} = S_{lm}(h,\eta) R_{lm}^{(j)}(h,\xi) e^{im\phi}, \tag{11}$$

where $h = kf = 2\pi f/\lambda$ is the spheroidal size parameter, S_{lm} is the spheroidal angle function, $R_{lm}^{(j)}$ is the spheroidal radial function [which may be of the first kind, $R_{lm}^{(1)}$, second kind, $R_{lm}^{(2)}$, third kind, $R_{lm}^{(3)} = R_{lm}^{(1)} + iR_{lm}^{(2)}$, or fourth kind, $R_{lm}^{(4)} = R_{lm}^{(1)} - iR_{lm}^{(2)}$], and the azimuthal dependence is contained in the complex exponential, $\exp(im\phi)$. The S_{lm} and $R_{lm}^{(j)}$ functions depend on h and are different for the prolate and oblate coordinate systems. l and m are integer indexes with $-\infty < m < \infty$ and, for a given value of m, $|m| < l < \infty$. A computer program for generating the associated eigenvalues, $\lambda_{lm}(h)$, and subroutines for generating the S_{lm} and $R_{lm}^{(j)}$ functions were written based on the procedures described in Flammer.[15,16] (See also, for example, Refs. 17–21.)

The incident acoustic pressure (assumed known) can be expressed in terms of a general eigenfunction expansion of the form

$$p^{(i)}(\xi,\eta,\phi) = \sum_{l,m} A_{lm} S_{lm}(h,\eta) R_{lm}^{(1)}(h,\xi) e^{im\phi}, \tag{12}$$

and the appropriate form of the eigenfunction expansion for the resultant scattered acoustic pressure (to be determined) is given by

$$p^{(s)}(\xi,\eta,\phi) = \sum_{l,m} a_{lm} S_{lm}(h,\eta) R_{lm}^{(3)}(h,\xi) e^{im\phi}. \tag{13}$$

In Eq. (12), only $R_{lm}^{(1)}$ is included in the expansion for the incident field since $R_{lm}^{(2)}$ becomes unbounded as ξ approaches the coordinate origin. In Eq. (13), $R_{lm}^{(3)}$ is used in the expansion for the scattered field since in the limit of large ξ this function corresponds to an outgoing traveling wave, appropriate for the scattered field solution. (In this same limit, the $R_{lm}^{(4)}$ function would correspond to an incoming traveling wave.) In practice, the double summation indexes are truncated such that $-M < m < M$ and $|m| < l < L$, where M and L are chosen sufficiently large for convergence of the solution.

The expansion coefficients for the incident field, A_{lm}, can be determined by evaluating the series expression for the incident field, as given in Eq. (12), at the surface of the spheroid, $\xi = \xi_0$, and then applying the orthogonality condition of the angle functions,

$$\int_0^{2\pi} \int_{-1}^{1} [S_{lm}(h,\eta) e^{im\phi}][S_{l'm'}(h,\eta) e^{-im'\phi}] d\eta\, d\phi = 2\pi N_{lm} \delta_{l,l'} \delta_{m,m'}; \tag{14}$$

thus

$$A_{lm} = \frac{1}{2\pi N_{lm} R_{lm}^{(1)}(h,\xi_0)} \int_0^{2\pi} \int_{-1}^{1} [p^{(i)}(\xi_0,\eta,\phi) \times S_{lm}(h,\eta) e^{-im\phi}] d\eta\, d\phi, \tag{15}$$

where N_{lm} is a constant that can be determined from the associated spheroidal function parameters, d_r^{lm} (see p. 22 of Flammer[15]),

$$N_{lm} = \int_{-1}^{1} [S_{lm}(h,\eta)]^2 d\eta = 2 \sum_{r=0',1'}^{\infty} \frac{(r+2m)!\,(d_r^{lm})^2}{(2r+2m+1)r!}, \tag{16}$$

where the summation is over even integers if $(l-m)$ is even and is over odd integers if $(l-m)$ is odd. The scattered field expansion coefficients, a_{lm}, can be directly related to the corresponding incident field coefficients, A_{lm}, by an application of the boundary condition that at the surface of the rigid spheroid the normal (ξ) component of the acoustic velocity must be zero. Since the acoustic velocity is directly proportional to the gradient of the acoustic pressure, and since the acoustic field consists of a sum of the incident and scattered fields, at $\xi = \xi_0$,

$$\frac{\partial p^{(s)}}{\partial \xi} + \frac{\partial p^{(i)}}{\partial \xi} = 0, \tag{17}$$

or, equivalently,

$$\frac{\partial p^{(s)}}{\partial \xi} = -\frac{\partial p^{(i)}}{\partial \xi}. \tag{18}$$

The series expansions for $p^{(s)}$ and $p^{(i)}$ are substituted into Eq. (18) and the series terms are equated mode-by-mode, which gives the following direct relationship between a_{lm} and A_{lm}:

$$a_{lm} = -\frac{R_{lm}^{(1)\prime}(h,\xi_0)}{R_{lm}^{(3)\prime}(h,\xi_0)} A_{lm}. \tag{19}$$

In summary, the procedure for determining the near-field acoustic pressure distribution for a rigid spheroid with a known incident field is as follows. The expansion coefficients of the incident field, A_{lm}, are calculated by numerical integration using Eq. (15). The expansion coefficients for the scattered field, a_{lm}, are then directly calculated from the A_{lm}

values using Eq. (19). Once the scattered field coefficients are determined, the series expansion of Eq. (13) can be used to determine the scattered field acoustic pressure anywhere external to the spheroid. The total acoustic pressure is equal to the sum of the scattered and incident fields.

III. NEAR-FIELD CALCULATIONS FOR THE PROLATE SPHEROID

Computer programs were written incorporating the theoretical development described in Sec. II. and systematic calculations investigating the effects of spheroid size, axis ratio, and the orientation and character of the incident field on the near-surface acoustic pressure distribution were performed. In this section, some results for the prolate spheroid geometry are presented. For plane wave incidence, the plane wave is assumed to propagate parallel to the *x*-*z* plane with an angle of incidence (θ_{dir}) measured relative to the *y*-*z* plane, as shown in Fig. 1. The corresponding expression for the incident acoustic pressure is thus

$$p^{(i)}(\mathbf{r}) = \exp[ih(x \sin\theta_{\text{dir}} + z \cos\theta_{\text{dir}})], \tag{20}$$

where, in this case, all acoustic pressures have been nondimensionalized relative to the amplitude of the incident plane wave. The expansion coefficients for a plane wave of the form given in Eq. (20) can, in fact, be determined analytically and, according to Eq. (5.3.3) in Ref. 15,

$$A_{lm} = i^l \frac{(2-\delta_{0m})}{N_{lm}} S_{lm}(h, \cos\theta_{\text{dir}}). \tag{21}$$

For plane wave incidence, the A_{lm} coefficients calculated using the numerical integration procedure of Eq. (15) were found (for both prolate and oblate geometries) to agree with the expected analytical values given in Eq. (21). For example, using ordinary trapezoidal rule numerical integration with an integration grid consisting of $4L$ steps in the η direction and $8M$ steps in the ϕ direction, the numerically calculated A_{lm} coefficients were typically found to agree with the theoretically predicted values given in Eq. (21) to better than one part in one million.

Figure 2 shows a surface grid plot of the calculated acoustic pressure in the *x*-*z* plane for a 45° angle of incidence plane wave incident upon a 2:1 axis ratio prolate spheroid with a spheroidal size parameter (h) of 10. The magnitude of the complex acoustic pressure is plotted using a 101 by 101 square grid extending from -2 to $+2$ in the x and z axis directions. (For the surface grid plots, the value of the acoustic pressure inside the spheroid is artificially set to a uniform value of 1.0.) Figure 3 provides a gray level visualization of the same data presented in the surface grid plot of Fig. 2. (For the gray level visualizations, the interior of the spheroid has been set to a uniform light gray). For the 2:1 axis ratio, $h=10$ prolate spheroid, solution convergence was obtained by using series limits of $M=12$ and $L=23$.

As can be observed in Figs. 2 and 3, the reflection of the incident plane wave from the lower surface (''incident side'') of the spheroid interferes with the oncoming incident plane wave to create a series of strong constructive and destructive interference bands. Near the upper surface (''shadow side'') of the spheroid there is a definite shielding effect except for a band of relatively high acoustic pressure that extends from the upper-right-hand surface in an outward direction that is approximately parallel with the propagation direction of the incident plane wave.

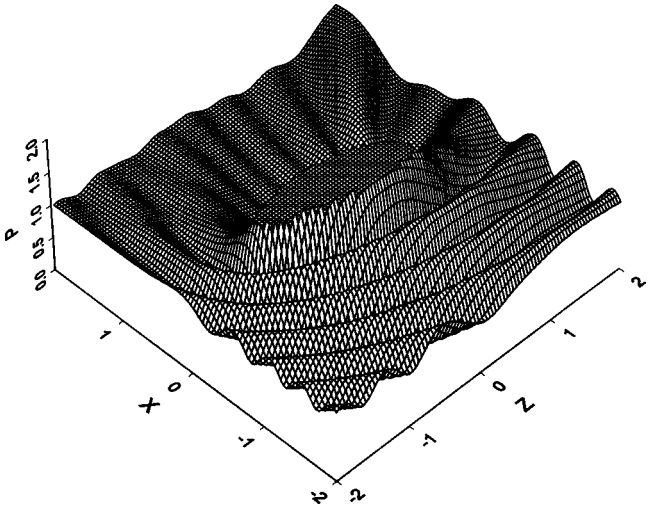


FIG. 2. A surface grid plot of the acoustic pressure in the *x*-*z* plane for a plane wave incident upon a 2:1 axis ratio prolate spheroid. Spheroid size parameter, $h=10$. Plane wave angle of incidence relative to the *y*-*z* plane, $\theta_{\text{dir}}=45°$.

By using the numerical integration procedure of Eq. (15), solutions can be obtained, not only for plane wave incidence, but for any known incident field. To demonstrate, a calculation was performed, for the same 2:1 axis ratio, $h=10$ prolate spheroid considered in Figs. 2 and 3, but now, in addition to the 45° angle of incidence plane wave, a spherical point source and a cylindrical line source (of the same frequency as the incident plane wave) are added. The spherical point source, located at $(x,y,z)=(2,0,2)$, was as-

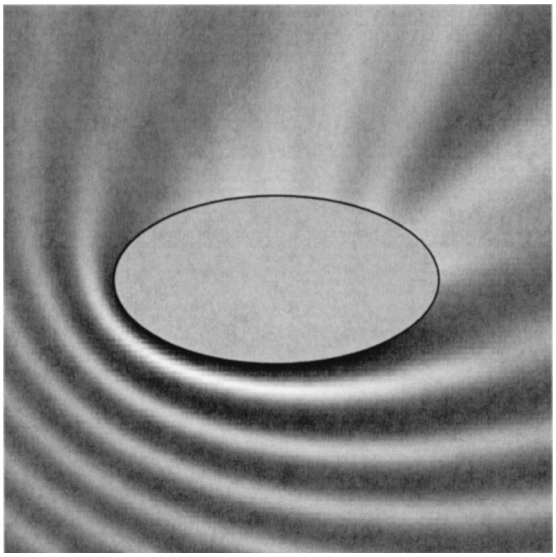


FIG. 3. Gray level visualization (white⇒low,black⇒high) of the acoustic pressure in the *x*-*z* plane for a plane wave incident upon a 2:1 axis ratio prolate spheroid. $h=10$ and $\theta_{\text{dir}}=45°$. The z axis (horizontal) extends from -2.0 on the left to $+2.0$ on the right. The x axis (vertical) extends from -2.0 at the bottom to $+2.0$ at the top. Maximum acoustic pressure, $p_{\max}=1.919$, and minimum acoustic pressure, $p_{\min}=0.360$.

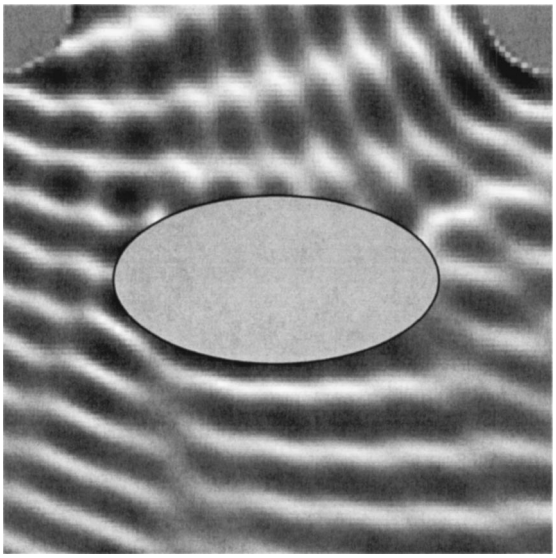


FIG. 4. Gray level visualization of the acoustic pressure in the *x*-*z* plane for a 2:1 axis ratio, $h=10$, prolate spheroid with a combined plane wave, spherical wave, and cylindrical wave sources. Plane wave: $p_{\mathrm{pl}}=1.0$, $\theta_{\mathrm{dir}}=45°$. Spherical wave: $p_{\mathrm{sph}}=0.5$, origin at (2,0,2). Cylindrical wave: $p_{\mathrm{cyl}}=3.0$, origin at (−2,0,2). The regions near the spherical and cylindrical sources have been blocked out. $p_{\max}=2.570$ and $p_{\min}=0.0199$.

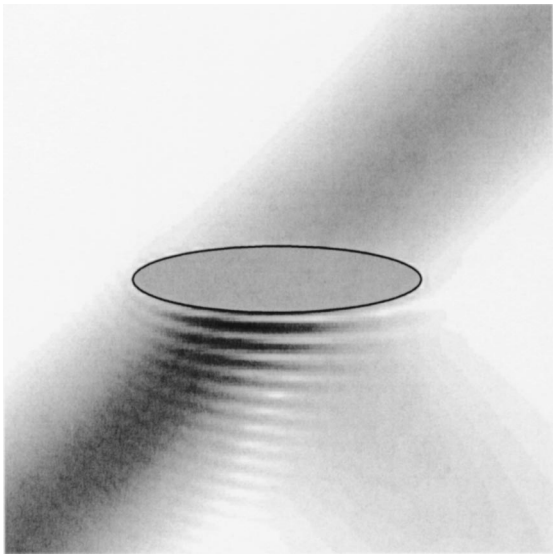


FIG. 5. Gray level visualization of the acoustic pressure in the *x*-*z* plane for a focused acoustic beam incident upon a 5:1 axis ratio prolate spheroid. $h=30$. Focused beam angle of incidence relative to the *y*-*z* plane, $\theta_{\mathrm{dir}}=45°$. Beam focused at the center of the spheroid ($x_0=0, y_0=0, z_0=0$) with a beam waist radius of $w_0=0.50$. $p_{\max}=1.939$ and $p_{\min}=0.000\,026$.

sumed to have a nominal amplitude one-half that of the incident plane wave, so

$$p_{\mathrm{sph}}(r')=(0.50/r')\exp(ihr'), \tag{22}$$

where r' is the radial distance from the point source. The cylindrical line source, located at $(x,y,z)=(-2,0,2)$ and extending in the *y*-axis direction, was assumed to have a nominal amplitude three times that of the incident plane wave, so that

$$p_{\mathrm{cyl}}(r'')=3.0[J_0(hr'')+iN_0(hr'')], \tag{23}$$

where J_0 and N_0 are, respectively, the zeroth-order cylindrical Bessel functions of the first and second kinds, and r'' is the plane radial distance from the line source. The complex pressure field resulting from all three sources (plane wave, spherical wave, and cylindrical wave) acting simultaneously was found to be identical to the sum of the complex pressure fields from the three sources acting individually, as expected from superposition.

Figure 4 provides a gray level visualization of the near-surface acoustic pressure for the combined plane wave, spherical wave, and cylindrical wave sources. (The regions immediately adjacent to the point and line sources have been intentionally blocked out so as to omit the very high acoustic pressures that occur near these source origins.) As can be observed in Fig. 4, the combined effects of the three sources interacting with the prolate spheroid produces a finer structured interference pattern in comparison with the plane wave only case given in Fig. 3.

The solution procedure is applicable for any known incident field (e.g., focused acoustic beam, acoustic scattering from an adjacent object, etc.). A gray level visualization of the acoustic pressure distribution for the arrangement of a focused acoustic beam incident upon a 5:1 axis ratio, $h=30$ spheroidal size parameter, prolate spheroid is provided in Fig. 5. The focused beam has a nondimensionalized beam waist radius (w_0) of 0.50, an angle of incidence of 45°, and is focused at the center of the spheroid, $(x_0,y_0,z_0)=(0,0,0)$. The analytical equations used to evaluate the pressure distribution of the Gaussian-profiled focused acoustic beam are given in the Appendix. These equations were derived using a procedure similar to that previously applied for the development of a fifth-order corrected electromagnetic focused Gaussian beam model.[22] For the 5:1 axis ratio, $h=30$ prolate spheroid, solution convergence was obtained by using series limits of $M=26$ and $L=51$.

As can be observed in Fig. 5, the incident acoustic beam strikes the lower surface (incident side) of the prolate spheroid and the weak reflected part of the beam interferes with the oncoming incident beam to create lines of constructive and destructive interference along the lower surface of the spheroid. Since the beam waist diameter ($2w=1.0$) is bigger than the minor axis diameter of the spheroid ($2b=0.408$), much of the incident acoustic beam is not reflected but instead finds its way around the spheroid and an intact, but decreased, amplitude, acoustic beam propagates away from the shadow side of the spheroid with the same direction of propagation as that of the incident beam.

Figure 6 is for the same conditions as Fig. 5, but now the focal point of the incident beam has been moved to the left tip of the spheroid, $(x_0,y_0,z_0)=(0,0,-1.0206)$. Figure 7 is similar, but the angle of incidence is changed to 0° (i.e., end-on incidence) and the focal point moved to the top edge of the spheroid, $(x_0,y_0,z_0)=(0.2041,0,0)$.

IV. NEAR-FIELD CALCULATIONS FOR THE OBLATE SPHEROID

In this section, some results for the oblate spheroid geometry are presented. The conditions of Figs. 8–13 corre-

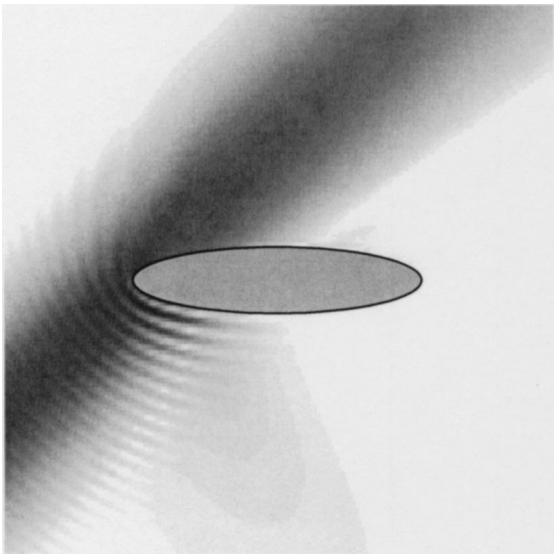


FIG. 6. Gray level visualization of the acoustic pressure in the *x*-*z* plane for a focused acoustic beam incident upon a 5:1 axis ratio prolate spheroid. $h=30$. A focused beam angle of incidence relative to the *y*-*z* plane, $\theta_{\mathrm{dir}}=45°$. The beam focused at the left tip of the spheroid ($x_0=0, y_0=0, z_0=-1.0206$) with a beam waist radius of $w_0=0.50$. $p_{\max}=1.711$ and $p_{\min}=0.000\,41$.

spond, respectively, with the conditions of Figs. 2–7, except now the spheroid is oblate, instead of prolate.

Figure 8 shows a surface grid plot of the calculated acoustic pressure in the *x*-*z* plane for a 45° angle of incidence plane wave incident upon a 2:1 axis ratio oblate spheroid with a spheroidal size parameter (h) of 10. Figure 9 provides a gray level visualization of the same data presented in the surface grid plot of Fig. 8. For the 2:1 axis ratio, $h=10$ oblate spheroid, solution convergence was obtained by using series limits of $M=15$ and $L=29$. As can be observed in

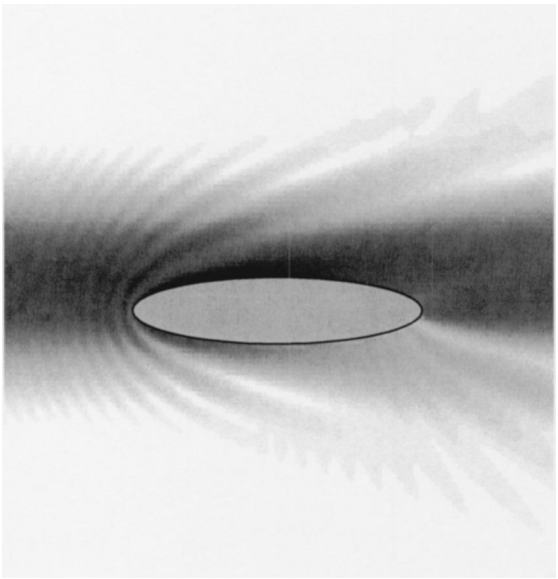


FIG. 7. Gray level visualization of the acoustic pressure in the *x*-*z* plane for a focused acoustic beam incident upon a 5:1 axis ratio prolate spheroid. $h=30$. Focused beam angle of incidence relative to the *y*-*z* plane, $\theta_{\mathrm{dir}}=0°$. The beam focused at the top edge of the spheroid ($x_0=0.2041, y_0=0, z_0=0$) with a beam waist radius of $w_0=0.50$. $p_{\max}=1.352$ and $p_{\min}=0.000\,50$.

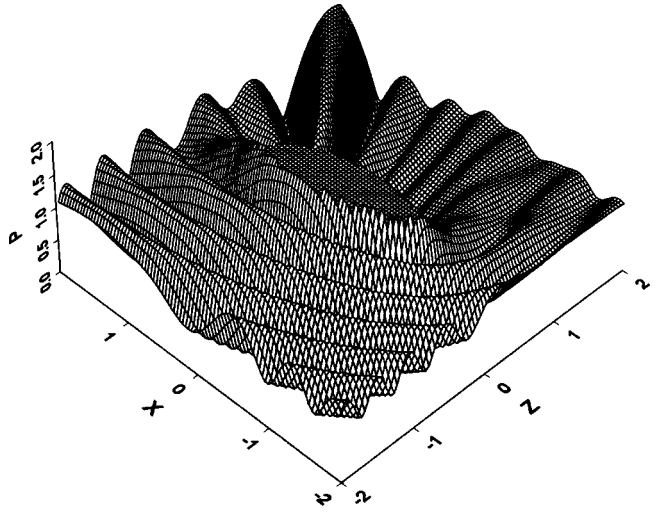


FIG. 8. Surface grid plot of the acoustic pressure in the *x*-*z* plane for a plane wave incident upon a 2:1 axis ratio oblate spheroid. $h=10$ and $\theta_{\mathrm{dir}}=45°$.

Figs. 8 and 9, the reflection of the incident plane wave from the left-hand surface (''incident side'') of the spheroid interferes with the oncoming incident plane wave to create a series of strong constructive and destructive interference bands. Near the right-hand surface (''shadow side'') of the spheroid there is some shielding and a band of relatively high acoustic pressure that extends from the upper-right-hand surface in an outward direction that is approximately parallel with the propagation direction of the incident plane wave.

Figure 10 provides a gray level visualization of the near-surface acoustic pressure for the combined plane wave, spherical wave, and cylindrical wave sources. Once again, as was observed for the prolate spheroid case, the combined effects of the three sources interacting with the oblate spheroid produces a finer structured interference pattern in comparison with the plane wave only case given in Fig. 9.

Figure 11 is for the same 5:1 axis ratio, $h=30$, on-center focused beam conditions as used in Fig. 5, but now an ob-

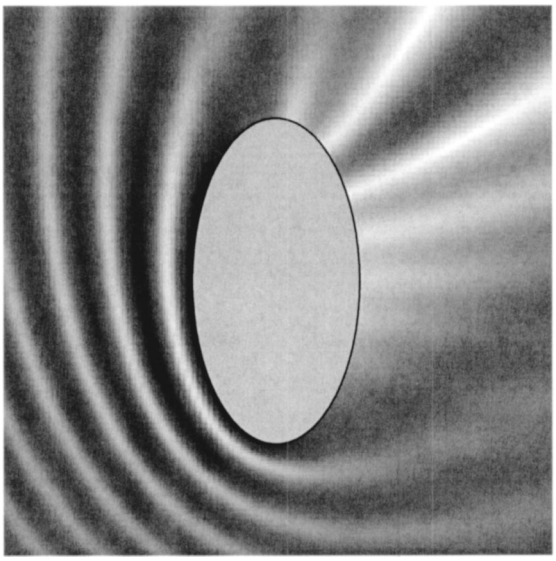


FIG. 9. Gray level visualization of the acoustic pressure in the *x*-*z* plane for a plane wave incident upon a 2:1 axis ratio oblate spheroid. $h=10$ and $\theta_{\mathrm{dir}}=45°$. $p_{\max}=1.981$ and $p_{\min}=0.0150$.

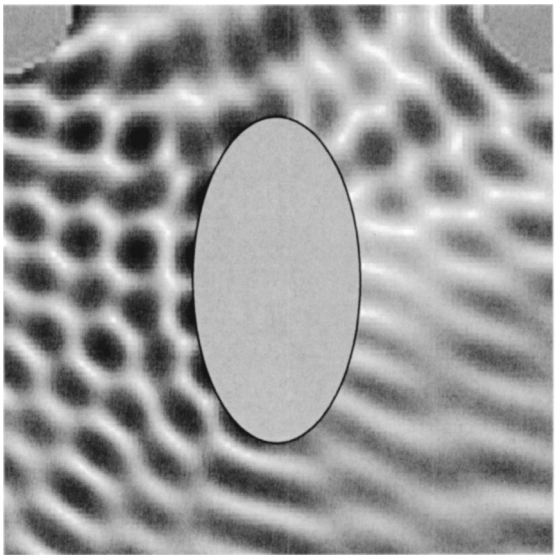


FIG. 10. Gray level visualization of the acoustic pressure in the *x*-*z* plane for a 2:1 axis ratio, $h=10$, oblate spheroid with combined plane wave, spherical wave, and cylindrical wave sources. Plane wave: $p_{\mathrm{pl}}=1.0$, $\theta_{\mathrm{dir}}=45°$. Spherical wave: $p_{\mathrm{sph}}=0.5$, origin at (2,0,2). Cylindrical wave: $p_{\mathrm{cyl}}=3.0$, origin at $(-2,0,2)$. The regions near the spherical and cylindrical sources have been blocked out. $p_{\max}=2.884$ and $p_{\min}=0.0070$.

late, instead of a prolate, spheroid is considered. For the 5:1 axis ratio, $h=30$ oblate spheroid, solution convergence was obtained by using series limits of $M=34$ and $L=55$. In the case of the oblate spheroid, since the beam waist diameter ($2w=1.0$) is smaller than the major axis diameter ($2a=2.041$), very little of the incident beam finds its way around the spheroid and almost all of the incident beam is reflected from the left-hand (incident-side) surface of the spheroid.

Figure 12 is for the same conditions as Fig. 11, but now the focal point of the incident beam has been moved to the bottom edge of the spheroid, $(x_0,y_0,z_0)=(-1.0206,0,0)$. Figure 13 is similar, but now the angle of incidence is changed to 90° (i.e., edge-on incidence) and the focal point moved to the right edge of the spheroid, $(x_0,y_0,z_0)=(0,0,0.2041)$.

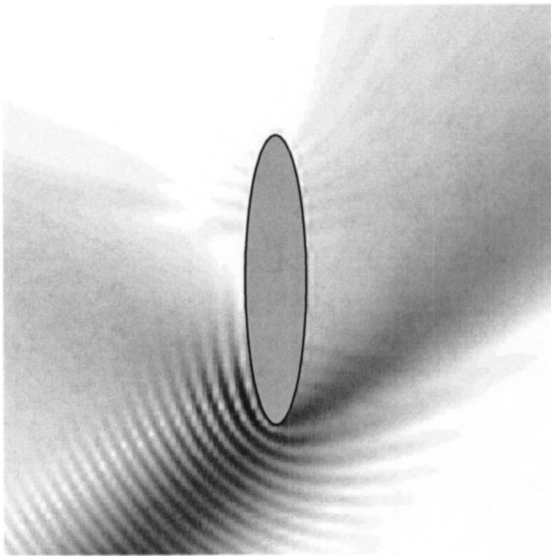


FIG. 12. Gray level visualization of the acoustic pressure in the *x*-*z* plane for a focused acoustic beam incident upon a 5:1 axis ratio oblate spheroid. $h=30$. A focused beam angle of incidence relative to the *y*-*z* plane, $\theta_{\mathrm{dir}}=45°$. The beam focused at the bottom edge of the spheroid ($x_0=-1.0206,y_0=0,z_0=0$) with a beam waist radius of $w_0=0.50$. $p_{\max}=1.935$ and $p_{\min}=0.0024$.

V. CONCLUSIONS

A general spheroidal coordinate separation-of-variables solution has been developed for the determination of the

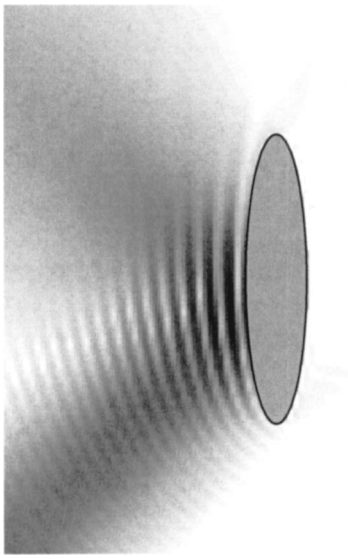


FIG. 11. Gray level visualization of the acoustic pressure in the *x*-*z* plane for a focused acoustic beam incident upon a 5:1 axis ratio oblate spheroid. $h=30$. The focused beam angle of incidence relative to the *y*-*z* plane, $\theta_{\mathrm{dir}}=45°$. The beam focused at the center of the spheroid ($x_0=0,y_0=0,z_0=0$) with a beam waist radius of $w_0=0.50$. $p_{\max}=1.999$ and $p_{\min}=0.000\,50$.

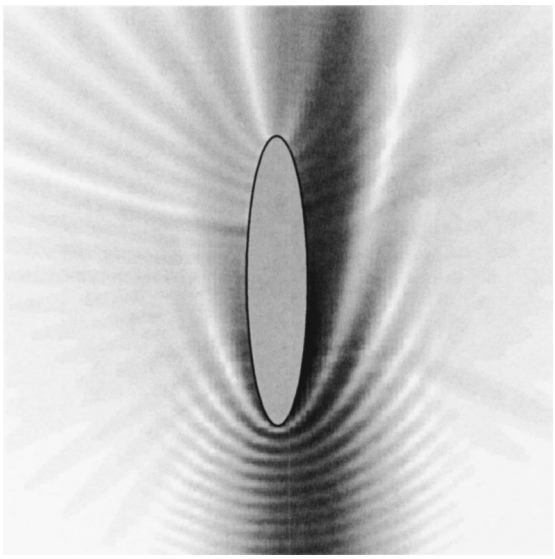


FIG. 13. Gray level visualization of the acoustic pressure in the *x*-*z* plane for a focused acoustic beam incident upon a 5:1 axis ratio oblate spheroid. $h=30$. Focused beam angle of incidence relative to the *y*-*z* plane, $\theta_{\mathrm{dir}}=90°$. The beam focused at the right edge of the spheroid ($x_0=0,y_0=0,z_0=0.2041$) with a beam waist radius of $w_0=0.50$. $p_{\max}=1.668$ and $p_{\min}=0.0061$.

acoustic pressure distribution near the surface of a rigid spheroid for a monofrequency incident acoustic field of arbitrary character. Calculated results have been presented, for both the prolate and oblate geometries, demonstrating the manner in which the orientation and character of the incident field can affect the subsequent near-surface acoustic pressure distribution. As an example of one possible application, a comparison of the near-field acoustic pressure distribution with the corresponding farfield scattering could be used to assist in developing insight into the physical origins of the angle-dependent structure exhibited in the farfield scattering pattern.

In general, higher axis ratio and/or higher size parameter objects require a larger number of series terms for convergence. The software programs used for the calculations presented here have an inherent upper spheroidal function limit of $M=40$ and $L=80$. With this limitation, calculations have been successfully performed for axis ratios up to 10:1 and spheroidal size parameters (h) up to 40.

APPENDIX: FOCUSED ACOUSTIC BEAM MODEL

A fifth-order corrected model for the acoustic pressure of a focused Gaussian-profiled beam with a beam waist radius of w_0 propagating in the $+z$-axis direction is as follows. An $\exp(-i\omega t)$ time dependence is assumed:

$$p(x,y,z)=p_r\left(1+s^2\frac{\psi_2}{\psi_0}+s^4\frac{\psi_4}{\psi_2}\right)iQ \times\exp(-i\rho^2Q)\exp(i\zeta/s^2), \tag{A1}$$

where

$$s=1/(hw_0),\quad \rho^2=(x^2+y^2),\quad \zeta=z/(hw_0),\quad Q=1/(i-2\zeta), \tag{A2}$$

and

$$\frac{\psi_2}{\psi_0}=(+2iQ+i\rho^4Q^3)$$

and

$$\frac{\psi_4}{\psi_2}=(-6Q^2-3\rho^4Q^4-2i\rho^6Q^5-\rho^8Q^6/2). \tag{A3}$$

[1] R. D. Spence and S. Granger, "The scattering of sound from a prolate spheroid," J. Acoust. Soc. Am. **23**, 701–706 (1951).

[2] A. Silbiger, "Scattering of sound by an elastic prolate spheroid," J. Acoust. Soc. Am. **35**, 564–570 (1963).

[3] T. B. A. Senior, "The scattering from acoustically hard and soft prolate spheroids for axial incidence," Can. J. Phys. **44**, 655–667 (1966).

[4] G. Lauchle, "Short-wavelength acoustic diffraction by prolate spheroids," J. Acoust. Soc. Am. **58**, 568–575 (1975).

[5] G. Lauchle, "Short-wavelength backscattering by a prolate spheroid," J. Acoust. Soc. Am. **58**, 576–580 (1975).

[6] W. K. Blake and G. A. Wilson, "Short-wavelength diffracted surface pressures on a rigid prolate spheroid," J. Acoust. Soc. Am. **61**, 1419–1426 (1977).

[7] A. Germon and G. C. Lauchle, "Axisymmetric scattering of spherical waves by a prolate spheroid," J. Acoust. Soc. Am. **65**, 1322–1327 (1979).

[8] R. H. Hackman, G. S. Sammelmann, K. L. Williams, and D. H. Trivett, "A reanalysis of the acoustic scattering from elastic spheroids," J. Acoust. Soc. Am. **83**, 1255–1266 (1988).

[9] G. S. Sammelmann, D. H. Trivett, and R. H. Hackman, "High-frequency scattering from rigid prolate spheroids," J. Acoust. Soc. Am. **83**, 46–54 (1988).

[10] V. K. Varadan, V. V. Varadan, L. R. Dragonette, and L. Flax, "Computation of rigid body scattering by prolate spheroids using the T-matrix approach," J. Acoust. Soc. Am. **71**, 22–25 (1982).

[11] L. Flax and L. R. Dragonette, "Analysis and computation of the acoustic scattering by an elastic prolate spheroid obtained from the T-matrix formulation," J. Acoust. Soc. Am. **71**, 1077–1082 (1982).

[12] G. C. Gaunaurd and M. F. Werby, "Interpretation of the three-dimensional sound fields scattered by submerged elastic shells and rigid spheroidal bodies," J. Acoust. Soc. Am. **84**, 673–680 (1988).

[13] Z. Ye, E. Hoskinson, R. K. Dewey, L. Ding, and D. M. Farmer, "A method for acoustic scattering by slender bodies. I. Theory and verification," J. Acoust. Soc. Am. **102**, 1964–1976 (1997).

[14] A. Maciulaitis, J. M. Seiner, and T. D. Norum, "Sound scattering by rigid oblate spheroids, with implication to pressure gradient microphones," NASA, 1976, TN D-8140, pp. 1–40.

[15] C. Flammer, *Spheroidal Wave Functions* (Stanford University Press, Stanford, CA, 1957).

[16] J. P. Barton, "Internal and near-surface electromagnetic fields for a spheroidal particle with arbitrary illumination," Appl. Opt. **34**, 5542–5551 (1995).

[17] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).

[18] J. A. Stratton, P. M. Morse, L. J. Chu, J. D. C. Little, and F. J. Corbato, *Spheroidal Wave Functions* (Wiley, New York, 1956).

[19] J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi, *Electromagnetic and Acoustic Scattering by Simple Shapes* (North-Holland, Amsterdam, 1969).

[20] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1970).

[21] S-J Zhang and J. Jin, *Computation of Special Functions* (Wiley, New York, 1996).

[22] J. P. Barton and D. R. Alexander, "Fifth-order corrected electromagnetic field components for a fundamental Gaussian beam," J. Appl. Phys. **66**, 2800–2802 (1989).

The role of the frozen surface approximation in small wave-height perturbation theory for moving surfaces

Richard S. Keiffer[a)]
Naval Research Laboratory, Code 7181, Stennis Space Center, Mississippi 39529-5004

Jorge C. Novarini
Planning Systems, Inc., Long Beach, Mississippi 39560-9702

Robert W. Scharstein
Electrical Engineering Department, University of Alabama, Tuscaloosa, Alabama 35487-0286

(Received 19 June 2002; revised 5 December 2002; accepted 17 December 2002)

In the standard development of the small wave-height approximation (SWHA) perturbation theory for scattering from moving rough surfaces [e.g., E. Y. Harper and F. M. Labianca, J. Acoust. Soc. Am. **58**, 349–364 (1975) and F. M. Labianca and E. Y. Harper, J. Acoust. Soc. Am. **62**, 1144–1157 (1977)] the necessity for any sort of frozen surface approximation is avoided by the replacement of the rough boundary by a flat (and static) boundary. In this paper, this seemingly fortuitous byproduct of the small wave-height approximation is examined and found to fail to fully agree with an analysis based on the kinematics of the problem. Specifically, the first order correction term from the standard perturbation approach predicts a scattered amplitude that depends of the source wave number, whereas the kinematics point to a scattered amplitude that depends on the scattered wave number. It is shown that a perturbation approach in which an explicit frozen surface approximation is made before the SWHA is invoked predicts (first order) scattered amplitudes that are in agreement with the analysis based on the kinematics. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1548153]

PACS numbers: 43.20.Fn, 43.20.Px, 43.30.Es, 43.30.Hw [DEC]

I. INTRODUCTION

The frozen surface approximation refers to a procedure in which the time-evolving surface involved in the scattering problem is replaced with a sequence of ''equivalent'' quasi-static surfaces. It is an indispensable part of many numerical schemes for solving the problem of scattering from a moving rough surface (e.g., Ref. 1), but not a topic usually associated with small wave-height approximation (SWHA) perturbation theory. As Labianca and Harper[2] discuss in their development of the perturbation method for moving sea surfaces, the small wave-height approximation replaces the original rough (moving) boundary with a flat (and obviously static) boundary. This *avoids* the necessity of making some sort of an explicit frozen surface approximation. Several of the most familiar treatments[3–5] of the SWHA perturbation theory for scattering from moving surfaces avoid the frozen surface approximation in this manner.

Our interest in this subject grew out of a study in which a computer code based on SWHA perturbation theory was developed and used to help benchmark a code that was based on a time-varying, impulse response technique. Those comparisons involved small amplitude (i.e., small compared to the shortest acoustic wavelengths of interest), pressure-release, moving sinusoidal surfaces and a collocated source and receiver positioned far from the scattering surface. The goal of this benchmarking exercise was to verify that the time-variant, impulse-response technique yielded the correct motion-induced changes in source and scattered frequency as well as changes in scattered amplitude for each of the energetically significant diffracted orders. In all comparisons, both models always precisely agreed on the locations (i.e., source and scattered frequencies) of any significant diffracted orders. However, the surface motion also changes the amplitudes of the (Doppler-shifted) diffracted orders and in this regard, the two models were in complete disagreement. SWHA perturbation theory predicted that the diffracted orders arriving at the receiver with frequencies higher than the source frequency (i.e., with positive Doppler shifts) also arrive with lower scattered amplitudes. Diffracted orders that are shifted toward lower received frequencies (negative Doppler shifts) arrive with larger amplitudes. For the time-variant impulse response method, the relationship between Doppler shift and scattered amplitude was exactly the opposite: diffracted orders that had positive Doppler shifts had larger scattered amplitudes, orders with negative Doppler shifts had smaller amplitudes. It was found that the magnitude of the model discrepancy grew as the speed of the surface increased. However, even for sinusoids moving much slower than the speed of sound the differences, while small (a few tenths of a dB for the intensities), were *not* negligible for benchmarking purposes. The divergence of these two predictions suggested to us that there was a difference in how the models were treating the kinematics of the problem.

That this was a subject worth a closer look was reinforced by the observations of Pourkaviani and Willemsen,[6] who, in a study that looked at the higher order perturbative corrections to the Doppler backscattering spectrum for ocean

a)Electronic mail: keiffer@nrlssc.navy.mil

gravity waves, pointed out that while it ''is natural within the static theory to organize terms (of the perturbation series) entirely on the basis of the order in perturbation theory, the Doppler theory forces an intrinsic reorganization depending on the regions of support in the frequency domain.'' Of practical importance, their study suggests that this intrinsic reorganization may lead to a breakdown in the SWHA theory at smaller values of roughness than in the static case.

The purpose of this paper is to point out the subtle but conceptually important role that the ''frozen-surface'' approximation plays in the SWHA perturbation method for moving surfaces. More precisely, this study examines what appear to be the unphysical consequences that occur if the frozen surface approximation is omitted. Our conclusions are interesting and surprising. We find that the perturbative approach in which an explicit frozen surface approximation is applied leads to results that are in agreement with fundamental kinematics of the problem. This is *not* true of the standard perturbation treatment in which the frozen surface approximation is avoided. A side by side development of the perturbation method with and without the frozen surface approximation (see Sec. II) is instructive in this regard. As we show, the two approaches yield first order correction terms that differ only in the amplitude they predict for the scattered field. In the section that follows, the kinematics of the problem is examined. There an independent derivation of the received spectral amplitude is presented that is based purely on the motion of the (Huygen) secondary sources on the surface. It is shown that the perturbation approach that incorporates an explicit frozen surface approximation yields scattered amplitudes that are in agreement with the kinematics of the problem.

Throughout the paper, the problem of scattering from a moving surface is formulated in the time domain, specifically for an impulsive point source and a solution in terms of the time-variant impulse response function is sought. This approach allows for a natural introduction of a single and multiple scattering description that, along with causality, makes it possible to identify ''what happened'' during the scattering process. It also allows us to address the kinematics of the problem in a methodical manner. Finally, working in the time domain avoids difficulties interpreting the frequency domain approach under dynamic, time varying conditions. Quoting Wetzel,[7] ''the sea is particularly appropriate for time domain modeling because it is a complex nonstationary surface characterized by an assortment of real and often identifiable, hydrodynamic scattering objects that can respond uniquely and individually to the fleeting touch of a passing impulse.''

To navigate between the time and frequency domains, and in particular to arrive at the response of the time-varying scattering system to the more familiar case of time harmonic excitation, some basic relations from time-variant linear filter theory are introduced. Under this viewpoint, the propagation of the impulse emitted at time $t=\tau$, from the source location $(\overline{r_0})$ to the surface, the scattering itself, and the propagation from the surface to the receiver $(\overline{r})$, are treated as a linear time-variant, space-variant filter. The complete specification of this filter is provided by the time-variant, space-variant, impulse response function $h(\overline{r},t;\overline{r_0},\tau)$. The received signal (i.e., the output of the filter), $p_{\text{out}}(\overline{r},t)$, due to input signal $p_{\text{in}}(\overline{r_0},\tau)$, is given by the superposition integral (see, e.g., Ref. 8)

$$p_{\text{out}}(\overline{r},t)=\int_{-\infty}^{\infty} p_{\text{in}}(\overline{r_0},\tau)h(\overline{r},t;\overline{r_0},\tau)d\tau. \tag{1}$$

If the input signal is time harmonic with frequency $\eta_0>0$, then the complex frequency spectrum of the output signal is

$$P_{\text{out}}(\overline{r},f)=\tfrac{1}{2}[B(\overline{r},f;\overline{r_0},\eta_0)+B(\overline{r},f;\overline{r_0},-\eta_0)]. \tag{2}$$

Here, $B(\overline{r},f;\overline{r_s},\eta_0)$ is called the bifrequency function. It is the Fourier transform of the time-variant transfer function

$$B(f;\eta)=\int_{-\infty}^{\infty} H(f;\tau)e^{-i2\pi\eta\tau}d\tau. \tag{3}$$

The time-variant transfer function is the Fourier transform of the impulse response function

$$H(f;\tau)=\int_{-\infty}^{\infty} h(t;\tau)e^{i2\pi ft}dt. \tag{4}$$

II. THE FROZEN SURFACE APPROXIMATION

As mentioned in the Introduction, the frozen surface approximation replaces the moving surface with a sequence of ''equivalent'' quasistatic surfaces. While this may seem like a familiar idea, the published literature offers few clear examples. Different quasistatic surfaces have been used, for example, a time sequence of instantaneous ''snapshots,''[1] but it has never been made clear in what sense these new surfaces are ''equivalent'' to the moving surface. In this section, causality, the kinematics of the surface insonification, and a single and multiple scattering description of the scattering process are used to motivate a practical and precise definition for the frozen surface approximation.

Consider a static point source, located at $\overline{r_0}=(x_0,y_0,z_0)$ above a moving rough surface $Z_S(x,y,t)$. The source emits an impulse (pressure) at time $t=\tau$ that propagates through a homogeneous medium (sound speed c) and then sweeps across the moving rough surface. The impulse response of the moving surface has, by definition, a single-scatter component that is due to those points on the interface that are directly (geometrically) insonified by the source and, acting as secondary sources, directly insonify the receiver. In this scheme, multiple scatter is a generic term that refers to scattered components that directly or indirectly result from the reinsonification of the surface by these *original* secondary sources.

At each point on the surface, $\overline{r}=(x,y,Z_S)$, the time of insonification by the source impulse, t_i, satisfies the equation

$$t_i=\tau+\frac{1}{c}\sqrt{(x_0-x)^2+(y_0-y)^2+(z_0-Z_S(x,y,t_i))^2}. \tag{5}$$

Causality requires that the scattered field on the rough interface be zero at all times prior to the arrival of the source impulse. Therefore, the time history of the surface at times prior to t_i has no effect on the scattered field. On the other hand, the single-scatter component of the impulse response

cannot depend on the surface height function at times later than t_i. The single-scatter response of a point on the surface must cease to exist the moment the source excitation ends. Thus, under the assumption that the single-scatter component dominates the impulse response, it makes sense to define the frozen surface using the source insonification time

$$\bar{Z}(x,y,\tau)=Z_S(x,y,t_i). \tag{6}$$

The frozen surface defined in this manner is equivalent to the moving surface in the sense that the single-scatter component of their respective impulse responses will be identical. The multiple scattering contributions to the impulse response functions, which can detect the surface evolution (or lack thereof) at times later than t_i, will in general differ. There are two factors that combine to govern the validity of this frozen surface approximation. The dependence of the impulse response function on multiple scattering determines the time span over which the surface evolution is effectively probed. The rate of the surface evolution determines how much the surface changes during this time span.

The equation for the insonification time can be determined numerically to machine precision via standard root finding techniques like the bisection method. To proceed analytically, assume that the source is far from all points on the surface of interest ($r_0 \gg r$), then the following approximation is valid:

$$\begin{aligned}|\overline{r_0}-\overline{r}| &\approx r_0-\overline{r}\cdot\hat{r}_0 \\ &= r_0-x\sin\theta_0\cos\phi_0-y\sin\theta_0\sin\phi_0-Z_s\cos\theta_0.\end{aligned} \tag{7}$$

Using the time of insonification of the flat $z=0$ surface as a zeroth order estimate, the method of successive substitutions can be used to provide an approximate expression for the frozen surface. For simplicity, we now assume that the surface is sinusoidal, $Z_S=a\cos(Kx-\Omega t)$, and restrict the source to the xz plane ($\phi_0=0,\pi$). The first set of estimates for the insonification time and the frozen surface are

$$t_i \approx t_i^{(0)}=t_0=\tau+\frac{1}{c}[r_0-x\sin\theta_0\cos\phi_0], \tag{8a}$$

$$\bar{Z}(x,y,\tau)\approx\bar{Z}^{(1)}(x,y,\tau)=a\cos[Kx-\Omega t_0]. \tag{8b}$$

Following the same procedure, the next higher order pair of estimates are

$$t_i \approx t_i^{(1)}=t_0-\frac{a\cos(Kx-\Omega t_0)\cos\theta_0}{c}, \tag{9a}$$

$$\begin{aligned}\bar{Z}(x,y,\tau) &\approx \bar{Z}^{(2)}(x,y,\tau) \\ &= a\cos[Kx-\Omega t_0+\beta\cos(Kx-\Omega t_0)],\end{aligned} \tag{9b}$$

with $\beta=(\Omega a/c)\cos\theta_0$. Note the differences between the frozen surface and the moving surface. The lowest order estimate for the frozen surface [Eq. (8b)] is sinusoidal but with a new wavelength $\Lambda'=2\pi/(K+\Omega\sin\theta_0\cos\phi_0/c)$. Sinusoids approaching the source appear (to the incident impulse) to have a shorter wavelength, sinusoids moving away from the source seem to have a longer wavelength. The next higher order estimate (9b) introduces the parameter β whose magnitude controls the phase modulation of the surface.

III. THE SMALL WAVE-HEIGHT PERTURBATION METHOD

Independent of the frozen surface approximation, the development of a perturbation solution that is valid in the limit of vanishing roughness begins with a power series expansion of the total field in surface height parameter a. As mentioned earlier, adopting the language of time-variant linear filter theory facilitates the interpretation of the two different perturbation developments. Accordingly, we write the expansion of the total pressure field at point $\overline{r}$ and time t (due to the source impulse emitted at time $t=\tau$ from position $\overline{r_0}$) as the expansion of the impulse response function

$$h(t;\tau)=h_0(t;\tau)+ah_1(t;\tau)+a^2h_2(t;\tau)+a^3h_3(t;\tau), \tag{10}$$

with unperturbed (flat surface) solution

$$h_0(t;\tau)=\frac{\delta\left(t-\tau-\dfrac{|\overline{r_0}-\overline{r}|}{c}\right)}{4\pi|\overline{r_0}-\overline{r}|}-\frac{\delta\left(t-\tau-\dfrac{|\overline{r_0}-2z_0\hat{z}-\overline{r}|}{c}\right)}{4\pi|\overline{r_s}-2z_0\hat{z}-\overline{r}|}. \tag{11}$$

A. Standard treatment

In the standard treatment (in which an explicit frozen surface approximation is avoided) the next step is to expand the impulse response function in a Taylor series about $z=0$ and evaluate the expansion on the surface. For pressure-release boundary conditions, this yields the equation

$$\begin{aligned}0=h(t;\tau)\big|_{z=0}&+Z_S(x,y,t)\left[\frac{\partial h(t;\tau)}{\partial z}\right]_{z=0} \\ &+\frac{Z_S^2(x,y,t)}{2}\left[\frac{\partial^2 h(t;\tau)}{\partial z^2}\right]_{z=0}+\frac{Z_S^3(x,y,t)}{6}\left[\frac{\partial^3 h(t;\tau)}{\partial z^3}\right]_{z=0} \\ &+\cdots.\end{aligned} \tag{12}$$

To proceed with the solution, the power series expression for the total impulse response [Eq. (10)] is substituted into the wave equation (not shown) and into the Taylor series expansion [Eq. (12)]. From these two new expressions, like powers of the expansion parameter a are equated. This yields a sequence of boundary value problems specifying the terms $a^n h_n(\overline{r},t)$ in Eq. (10). Our initial interest lies with the first order correction to the unperturbed impulse response. Using $Z_S=a\cos(Kx-\Omega t)$ with $K>0$, the first order correction to the unperturbed impulse response has the form

$$ah_1(t;\tau)\big|_{z=0}=-a\cos(Kx-\Omega t)\frac{\partial}{\partial z}h_0(\overline{r},t;\overline{r_0},\tau)\big|_{z=0}. \tag{13}$$

Note that evaluating the z derivative of unperturbed impulse response on the $z=0$ interface introduces the generalized function $\delta'(t-t_0)$. Here t_0 is the time at which the impulse, fired at time $t=\tau$, reaches the flat $z=0$ interface at horizontal location (x,y), and the prime indicates $\partial/\partial t$. Making use of operational properties of the derivative of the Dirac delta

function[9] allows the first order correction term to be written as

$$ah_1(t;\tau)|_{z=0}=\frac{a\cos\theta_0}{2\pi cr_0}[\cos(Kx-\Omega t_0)\,\delta'(t-t_0) - \Omega\sin(Kx-\Omega t_0)\,\delta(t-t_0)], \quad (14)$$

where the time of insonification for the flat surface (t_0) is given by Eq. (8a). Note that although no explicit frozen surface approximation was made in the development of Eq. (14), the frozen surface approximation has implicitly been invoked. This can be seen from the fact that, at each point on the surface, the scattered impulse response in Eq. (14) depends on the surface evolution at only one time, $t=t_0$.

The time-variant transfer function (corresponding to the first order perturbation correction) results after the Fourier transform indicated in Eq. (4) is performed on Eq. (14),

$$aH_1(f;\tau)|_{z=0}=\frac{-a\exp[i2\pi ft_0]\cos\theta_0}{2\pi cr_0} \times[i2\pi f\cos(Kx-\Omega t_0)-\Omega\sin(Kx-\Omega t_0)]. \quad (15)$$

The bifrequency function is the Fourier transform of the transfer function and for the purposes of this paper, it is unnecessary to continue the development any farther. Recall that the acoustic response of the moving surface due to time harmonic excitation is rather simply related to the bifrequency function [Eq. (2)]. In addition, the salient features of the problem do not require that the scattered field on the flat surface be propagated into the space above the surface. In the absence of an explicit frozen surface approximation, the standard perturbation development yields (for a source frequency $\eta_0>0$) the first order correction to the scattered bifrequency function

$$aB_1(f;\eta_0)|_{z=0}=\frac{ia\eta_0\cos\theta_0}{2cr_0}\exp\left[i\frac{2\pi\eta_0}{c}r_0\right] \times\left[\exp\left[-i\left(\frac{2\pi\eta_0}{c}\sin\theta_0\cos\phi_0-K\right)x\right] \times\delta\left(f-\eta_0-\frac{\Omega}{2\pi}\right) +\exp\left[-i\left(\frac{2\pi\eta_0}{c}\sin\theta_0\cos\phi_0+K\right)x\right] \times\delta\left(f-\eta_0+\frac{\Omega}{2\pi}\right)\right]. \quad (16)$$

The first order correction to the scattered bifrequency function consists of two plane wave components, one Doppler shifted toward higher frequencies and one shifted toward lower frequencies. The critical feature to note is that both components have amplitudes that depend on the nondimensional parameter k_0a, where $k_0=2\pi\eta_0/c$ is the source wave number. Higher order corrections depend on the scattered wave number so that, for example, the *n*th order correction to the scattered amplitude of a Doppler-shifted diffracted order has an amplitude that is proportional to $(k_0a)(ka)^{n-1}$ with $k=2\pi f/c$ being the scattered wave number.

B. Perturbation approach coupled with an explicit frozen surface approximation

If the frozen surface approximation is applied, the development is identical to the standard theory except that the surface $Z_s(x,y,t)$ is replaced by the equivalent quasistatic surface $\bar{Z}_s(x,y,\tau)$. Corresponding to Eq. (12), but using $\tilde{h}(t;\tau)$ to distinguish the frozen surface impulse response function, is the Taylor series expansion of the boundary conditions,

$$0=\tilde{h}(t;\tau)|_{z=0}+\bar{Z}_S(x,y,\tau)\left[\frac{\partial\tilde{h}(t;\tau)}{\partial z}\right]_{z=0} +\frac{\bar{Z}_S^2(x,y,\tau)}{2}\left[\frac{\partial^2\tilde{h}(t;\tau)}{\partial z^2}\right]_{z=0}+\frac{\bar{Z}_S^3(x,y,\tau)}{6}\left[\frac{\partial^3\tilde{h}(t;\tau)}{\partial z^3}\right]_{z=0} +\cdots. \quad (17)$$

After the frozen surface approximation has been applied, the equation for the first order correction term [corresponding to Eq. (13)] is

$$a\tilde{h}_1(t;\tau)|_{z=0}=-a\bar{Z}_S(x,y,\tau)\frac{\partial}{\partial z}h_0(\bar{r},t,\tau)|_{z=0}. \quad (18)$$

Using Eqs. (8a) and (8b) to approximate the insonification time and the frozen surface, the first order correction to the time-variant transfer function becomes

$$a\tilde{H}_1(f;\tau)|_{z=0}=\frac{iaf\cos\theta_0}{cr_0}\cos(Kx-\Omega t_0)\exp[i2\pi ft_0]. \quad (19)$$

A comparison between Eqs. (15) and (19) is interesting because it reveals (at the level of the time-variant transfer function) the effect that separating the surface evolution from the time dependence of the incident field has on the scattered field. Note that the transfer functions are identical except for the extra term in Eq. (15) that comes from the time derivative of the surface height function. The bifrequency function that follows from Eq. (19) (for a source frequency $\eta_0>0$) is

$$a\tilde{B}_1(f;\eta_0)|_{z=0}=\frac{iaf\cos\theta_0}{2cr_0}\exp\left[i\frac{2\pi\eta_0}{c}r_0\right] \times\left[\exp\left[-i\left(\frac{2\pi\eta_0}{c}\sin\theta_0\cos\phi_0-K\right)x\right] \times\delta\left(f-\eta_0-\frac{\Omega}{2\pi}\right) +\exp\left[-i\left(\frac{2\pi\eta_0}{c}\sin\theta_0\cos\phi_0+K\right)x\right] \times\delta\left(f-\eta_0+\frac{\Omega}{2\pi}\right)\right]. \quad (20)$$

Rewriting this in a slightly different form for a more direct comparison with Eq. (16),

$$a\tilde{B}_1(f;\eta_0)|_{z=0}=\frac{ia\eta_0\cos\theta_0}{2cr_0}\exp\left[i\frac{2\pi\eta_0}{c}r_0\right]\left[\left(1+\frac{\Omega}{2\pi\eta_0}\right)\times\exp\left[-i\left(\frac{2\pi\eta_0}{c}\sin\theta_0\cos\phi_0-K\right)x\right]\times\delta\left(f-\eta_0-\frac{\Omega}{2\pi}\right)+\left(1-\frac{\Omega}{2\pi\eta_0}\right)\times\exp\left[-i\left(\frac{2\pi\eta_0}{c}\sin\theta_0\cos\phi_0+K\right)x\right]\times\delta\left(f-\eta_0+\frac{\Omega}{2\pi}\right)\right]. \quad (21)$$

It can be seen that the application of an explicit frozen surface approximation (and here we have used a zeroth order estimate for the frozen surface), results in a bifrequency function that contains the same two components found in the standard approach [see Eq. (16)]. The difference in the two developments is in the scattered amplitudes. With the explicit frozen surface approximation, the scattering amplitudes depend on the nondimensional parameter ka, not k_0a as in the standard approach. Thus, if the Doppler shift is toward higher frequencies the scattering amplitude increases, if the Doppler shift is toward lower frequencies then the scattering amplitude decreases. As pointed out earlier, the higher order corrections depend on the scattered wave number; the nth order correction to the scattered amplitude of a diffracted order has an amplitude that is proportional to $(ka)^n$.

For illustrative purposes, consider the form of the first order correction to the scattered bifrequency function if the higher order estimates [Eqs. (9a) and (9b)] are used for the time of insonification, t_i, and frozen surface, $\bar{Z}(x,y,\tau)$,

$$a\tilde{B}_1(f;\eta_0)|_{z=0}=\frac{ia\eta_0\cos\theta_i}{2cr_s}\exp\left[i\frac{2\pi\eta_0}{c}r_s\right]\left[\sum_{m=0}^{\infty}\left(1+\frac{m\Omega}{2\pi\eta_0}\right)(J_{m-1}(\beta)-J_{m+1}(-\beta))\exp\left[-i\left(\frac{2\pi\eta_0}{c}\sin\theta_0\cos\phi_0-mK\right)x\right]\times\delta\left(\eta_0-f+\frac{m\Omega}{2\pi}\right)+\sum_{m=0}^{\infty}\left(1-\frac{m\Omega}{2\pi\eta_0}\right)(J_{m-1}(-\beta)-J_{m+1}(\beta))\exp\left[-i\left(\frac{2\pi\eta_0}{c}\sin\theta_0\cos\phi_0+mK\right)x\right]\delta\left(\eta_0-f-\frac{m\Omega}{2\pi}\right)\right]. \quad (22)$$

The frozen surface given by Eq. (9b) included a term that modulated the frequency of the surface. The effect of this modulation on the scattered spectrum is to generate an infinite number of side bands. These additional components arrive at the receiver with frequencies that differ from the source by multiples of the modulation frequency $(\Omega/2\pi)$. The modulation index, $\beta=(\Omega a/c)\cos\theta_0$, controls the number of side bands having significant spectral energy. Since by assumption, $ka<1$, this modulation effect will be very small except for those cases where the value of ka approaches 1 and the sinusoidal surface moves at a significant fraction of the speed of sound.

IV. ANALYSIS OF THE KINEMATICS

An understanding of the impact of the kinematics on the scattered amplitude can be obtained without actually calculating the scattering. For a uniformly moving rough surface, the relationship between the source and received (scattered) frequency can be deduced by considering the time variation in the impulse response of a single scattering feature. For clarity, label this scattering feature ξ. Suppose $\bar{Z}_S(x_\xi,0)$ is the surface height that is detected [Eq. (6)] at horizontal location (x_ξ,y_ξ) by the source impulse emitted at $t=\tau=0$. The vector that locates this surface feature is

$$\bar{r}_{\xi_0}=x_\xi\hat{x}+y_\xi\hat{y}+\bar{Z}_S(x_\xi,0)\hat{z}. \quad (23)$$

An impulse emitted at a later time, $t=\tau$, will detect this *same* surface feature at a new horizontal location

$$\bar{r}_\xi=\bar{r}_{\xi_0}+u\tau\left(\frac{c}{c+u\sin\theta_0\cos\phi_0}\right)\hat{x}. \quad (24)$$

Here $u\hat{x}$ is the surface velocity. The single-scatter component of the time-variant impulse response for this moving surface feature can be written as

$$h_\xi(t,\tau)=\sigma_\xi\delta\left(t-\tau-\frac{|\bar{r}_0-\bar{r}''|+|\bar{r}-\bar{r}''|}{c}\right). \quad (25)$$

Here σ_ξ is an unknown amplitude that varies with scattering geometry and surface feature. It also accounts for the spreading losses from source to the scattering feature and from scattering feature to the receiver. Under the assumption that $r_0\gg r_\xi$ and $r\gg r_\xi$, σ_ξ can be considered time invariant. After some algebra, the moving features' impulse response becomes

$$h_\xi(t,\tau)=\sigma_\xi\delta(t-q\tau-T_0), \quad (26)$$

where

$$q=\frac{c-u\sin\theta\cos\phi}{c+u\sin\theta_0\cos\phi_0} \quad (27)$$

and

$$T_0 = \frac{r_0 + r - (\hat{r}_0 + \hat{r}) \cdot \overline{r}_{\xi_0}}{c}. \tag{28}$$

In this form it easy to perform the two sequential Fourier transforms that yield the bifrequency function

$$B_\xi(f, \eta) = \frac{\sigma_\xi}{q} \exp[i2\pi f T_0] \delta\left(f - \frac{\eta}{q} \right). \tag{29}$$

At this point, it can be noted that the amplitude of this bifrequency function increases when the Doppler shift is toward frequencies that are higher than the source (frequency) and decreases when the Doppler shift is toward lower frequencies. This is qualitatively the same change to the received amplitude that was displayed by [Eq. (21)], the bifrequency function derived from the frozen surface SWHA perturbation theory. Since every point on the surface is in uniform motion, this scaling effect applies to the impulse response of the entire surface.

To make the correspondence of the above scaling factor more explicitly similar to that of Eq. (21), consider a periodic surface and add the conditions on the source and received frequencies for constructive interference into the receiver direction. Two identical points on the surface, separated by one surface wavelength, will constructively interfere at the receiver for frequency f provided that, $f = n/\Delta T$. Here n is an integer and ΔT is the difference in the time of flight from the two surface points to the receiver. For a receiver far from all points of interest on the surface, this becomes

$$f_n = \frac{n(c + u \sin\theta_0 \cos\phi_0)}{\Lambda(\sin\theta_0 \cos\phi_0 + \sin\theta \cos\phi)}. \tag{30}$$

Using the connection between the source and receiver frequencies established by the delta function in Eq. (29), allows for the similar specification of the conditions on the source frequency for constructive interference at the receiver

$$\eta_n = \frac{n(c - u \sin\theta \cos\phi)}{\Lambda(\sin\theta_0 \cos\phi_0 + \sin\theta \cos\phi)}. \tag{31}$$

Note, by convention, restricting the interest to positive frequencies with $(\sin\theta_0 \cos\phi_0 + \sin\theta \cos\phi) > 0$ means choosing $n > 0$. The delta function in Eq. (29) means that $1/q = f/\eta$. Injecting the requirements for constructive interference changes this to $f_n/\eta_n = 1 + (nu/\Lambda\eta_n)$. Finally, identifying $u/\Lambda = \Omega/2\pi$ yields the bifrequency

$$B_\xi(f, \eta_n) = \left(1 + \frac{n\Omega}{2\pi\eta_n} \right) \sigma_\xi \exp[i2\pi f T_0] \delta\left(f - \eta_n - \frac{n\Omega}{2\pi} \right). \tag{32}$$

For $n = 1$, this equation has exactly the frequency dependent scaling factor, $1 + (\Omega/2\pi\eta_m)$, as found in Eq. (21).

V. CONCLUDING REMARKS

In the preceding section, an analysis of the kinematics of the moving sinusoid scattering problem was shown to account for a change in the received frequency (Doppler shift) and a scaling of the received amplitude that depends on the Doppler shift. Both the kinematically derived Doppler shift and amplitude scaling effect exactly agree with the first order correction from the SWHA perturbation theory that explicitly invoked the frozen surface approximation. On the other hand, the standard development of the perturbation theory, which appears to avoid the frozen surface approximation, accounts for the Doppler shift but not the amplitude scaling effect.

Initially, it may not be obvious which of the two perturbation solutions makes the most physical sense. From the side by side development of the two approaches, it is clear that the frozen surface approximation forces the separate treatment of the kinematics and scattering (dynamic) aspects of the problem. In the standard development, the kinematics are buried in the perturbation series and become part of the scattering problem. While it may be that the additional refinements to the scattered field that are due purely to the kinematics will emerge from the standard treatment with the incorporation of higher order terms, some differences between the two approaches will persist. On the one hand, this is apparent from the fact that all higher correction terms are derived from the first order correction term. On the other hand, the perturbation calculation is limited in a practice to the first few corrections terms and the desired spectral information may not have emerged adequately with only a few terms. For example, for a ''fast'' sinusoid the first order correction term of the scattered field [Eq. (22)] displays a side band structure that can contain a significant fraction of the scattered energy. The frozen surface approximation technique makes some of this spectral information available with the first order correction term. With the standard perturbation approach, information about the existence and significance of these side bands emerges (if at all) with each additional higher order (in surface height) correction term.

The main purpose of this paper is to point out that by addressing the kinematics separately from the SWHA perturbation development, one can obtain (for each order of the perturbative series) a more accurate accounting of the scattered spectrum than is possible via the standard perturbation development. The case of scattering from a sinusoidal surface was considered largely for pedagogical purposes. It should not be assumed that the specific differences between the two perturbation approaches cited herein translate entirely to more complex surfaces. For these moving multiscale surfaces, the advantages of the perturbation method that uses an explicit frozen surface approximation may be greater.

ACKNOWLEDGMENTS

This work has been supported by the Office of Naval Research, Program Element No. 61153N-32 and by grants of computer time at the DoD High Performance Computing Shared Resource Center located at Stennis Space Center, MS. This document has been reviewed and is approved for public release.

[1] C. L. Rino, T. L. Crystal, A. K. Koide, H. D. Ngoc, and H. Guthart, ''Numerical simulation of backscatter from linear and nonlinear ocean surface realizations,'' Radio Sci. **26**, 51–71 (1991).

[2] F. M. Labianca and E. Y. Harper, ''Connection between various small-waveheight solutions of the problem of scattering from the ocean surface,'' J. Acoust. Soc. Am. **62**, 1144–1157 (1977).

[3] E. Y. Harper and F. M. Labianca, ''Scattering of sound from a point source by a rough surface progressing over an isovelocity ocean,'' J. Acoust. Soc. Am. **58**, 349–364 (1975).
[4] F. G. Bass and I. M. Fuks, *Scattering from Statistically Irregular Surfaces* (National Technical Information Service, U.S. Dept. of Commerce, Springfield, VA, 1975), Chap. 5.
[5] L. M. Brekhovskikh and Y. P. Lysanov, *Fundamentals of Ocean Acoustics*, 2nd ed. (Spinger-Verlag, Berlin, Germany, 1991), pp. 183–300.
[6] M. Pourkaviani and J. F. Willemsen, ''Perturbative corrections to acoustic Doppler backscattering I: Ocean gravity waves,'' J. Acoust. Soc. Am. **90**, 426–432 (1991).
[7] L. B. Wetzel, ''A time domain model for sea scatter,'' Radio Sci. **28**, 139–150 (1993).
[8] L. J. Ziomek, *Underwater Acoustics: A Linear Systems Theory Approach* (Academic, Orlando, FL, 1985), Chap. 2.
[9] R. N. Bracewell, *The Fourier Transform and Its Applications* (McGraw-Hill, New York, 1978), pp. 209–211.

Imaging in the presence of grain noise using the decomposition of the time reversal operator

E. Kerbrat, C. Prada, D. Cassereau, and M. Fink
Laboratoire Ondes et Acoustique, ESPCI, Université Paris VII, UMR CNRS 7587, 10 rue Vauquelin, 75005 Paris, France

(Received 15 April 2002; revised 25 November 2002; accepted 17 December 2002)

In this paper, we are interested in detecting and imaging defects in samples of cylindrical geometry with large speckle noise due to the microstructure. The time reversal process is an appropriate technique for detecting flaws in such heterogeneous media as titanium billets. Furthermore, time reversal can be iterated to select the defect with the strongest reflectivity and to reduce the contribution of speckle noise. The DORT (the French acronym for Decomposition of the Time Reversal Operator) method derives from the mathematical analysis of the time reversal process. This detection technique allows the determination of a set of signals to be applied to the transducers in order to focus on each defect separately. In this paper, we compare three immersion techniques on a titanium sample, standard transmit/receive focusing, the time reversal mirror (TRM), and the DORT method. We compare the sensitivity of these three techniques, especially the sensitivity to a poor alignment of the array with the front face of the sample. Then we show how images of the sample can be obtained with the TRM and the DORT method using backpropagation algorithm. [DOI: 10.1121/1.1548156]

PACS numbers: 43.20.Gp, 43.20.Bi [DEC]

I. INTRODUCTION

Most applications in nondestructive testing require the focusing of ultrasonic waves inside the investigated medium in order to detect minor defects. The detection of weak flaws is especially difficult when the inspected object is heterogeneous[1] and has complex geometry.[2] The weak echo from the defect can be masked by the speckle due to the microstructure and the complex geometry significantly degrades the propagation of ultrasound in the area of interest. In many applications, the solid sample is immersed in water and one or several transducers are moved to scan the area of interest. Different approaches have been extensively studied in order to produce an ultrasonic focused beam through complex liquid–solid interfaces. In the simplest approach, the beam focusing is achieved with one transducer whose geometry is matched to the liquid–solid interface and the desired focal point.[3] Thus, the focusing process is improved allowing a better signal-to-noise ratio and also a greater resolution assisting the separation of close defects and flaw sizing. However, the inspection of thick samples requires the use of several transducers of different geometries in order to scan through the whole sample. This step-by-step scanning is also extremely time consuming. Last but not least, a misadjustment of the focused transducer significantly degrades the detection quality. This bad positioning is frequently observed, especially for large samples when the focused transducer is translated over a long distance.

A second technique based on the use of multielement transducer arrays gives a greater flexibility.[4] Transducer arrays are connected to a set of electronic time delays whose values are adapted to focus in both the transmit and receive modes. The combination of transmit and receive focusing allows the achievement of high resolution.

Another technique, called SAFT[5,6] (Synthetic Aperture and Focusing Technique) is based on the coherent summation of multiple data obtained from the sample, from a large number of transducers positions.

However, all these techniques require some prior knowledge of the geometry and of the acoustics properties of the solid sample.

Here, we are interested in detecting and imaging the defect in titanium billets used in the aerospace industry. Such samples are heterogeneous media and the polycrystalline microstructure yields a strong scattering noise,[7] which can hide the echo from a defect.

To solve the detection problems due to the heterogeneous structure and the geometry of the sample, another technique has been developed in our laboratory. Relying on the time reversal invariance of the acoustic propagation, this technique uses a time reversal mirror[8,9] (TRM). Such a TRM can convert a divergent wave, scattered by a defect, into a convergent wave focused on the defect, by emitting a time reversed version of the received wave field. It is a self-focusing technique[10] that compensates for distortions of the ultrasonic wave due to the geometry of the liquid–solid interface and to the microstructure of the sample.[11]

When the sample contains only one defect, after a time reversal (TR) process, an optimal focusing is obtained on the flaw, without the need for any assumptions concerning the geometry of the sample and the ultrasonic celerity. In the presence of speckle noise, the iteration of the TR operation allows the separation of the echo of a weak, deep defect from the speckle noise.[12] If the sample contains several defects, several iterations of the TR process leads to focusing on the strongest flaw.[12,13]

Parallel to the implementation of a detection procedure using time reversal mirrors, we have developed another detection technique called the DORT (the French acronym for

Decomposition of the Time Reversal Operator) method. This method is based on the decomposition of the time reversal operator that was introduced to describe the iterative time reversal process.[13] Similar to the time reversal process only rough assumptions need to be made on the ultrasound velocity and the geometry of the medium.

The DORT method has been used to separate the responses from different scatterers in a wave guide,[14] through inhomogeneous media,[15] and also in thin plates using Lamb waves.[16] It has also been applied to the separation and characterization of Lamb waves circumnavigating a shell.[17]

The goal of this paper is to show that the DORT method can be efficiently applied in nondestructive testing. In particular, we show that the DORT method can be used to enhance and separate the echo of a weak scatterer from speckle noise. In order to evaluate the potentials of this technique, we compare it to classical focusing and to the time reversal process.

The DORT method and the time reversal process are then combined with numerical backpropagation of the data in order to reconstruct an image of the medium. This backpropagation is calculated using the simulation code PASS that takes into account the geometry of the transducers as well as the solid–liquid interface.[18] Thus, it requires some prior knowledge of the sample.

II. EXPERIMENTAL SETUP

The three detection techniques have been compared using the same setup: a cylindrical titanium sample with defects, and a transducer array connected to a multichannel electronic system including programmable transmitters.

A. The transducer array and the electronic prototype

Experiments are realized with a ''Fermat surface'' transducer array whose geometry is adapted to make all the propagation times equal from the transducer surface to a point at 140 mm depth in the inspected sample, thus providing optimum focusing at this point. The array of transducers is prefocused at 140 mm in the titanium billet of 250 mm diameter with a 55 mm deep water column. This array is made of 121 transducers working at 5 MHz, with a bandwidth equal to 70%. The transducer elements, having the same area, are distributed according to a six annuli structure of, respectively, 1, 8, 16, 24, 32, and 40 elements (Fig. 1).

The 121 elements are connected to an electronic system allowing transmission and acquisition of the experimental data for the three detection tests at a sampling rate of 40 MHz. The transmission module is made of 121 programmable transmitters. This enables us to emit and to record any kind of temporal signal on each transducer of the array.

B. The inspected solid sample

The inspected medium is a Ti6-4 titanium alloy cylindrical sample (Fig. 2), immersed in water, that contains three flat bottom holes (FBH). These defects of diameters 0.8, 0.4, and 0.5 mm are located at a depth of 140 mm and spaced 15 mm apart. The longitudinal wavelength in the titanium sample is 1.2 mm at 5 MHz and the transverse resolution at 140 mm inside the material is 3.3 mm at -6 dB. Thus, the defects are well resolved and can be considered as points.

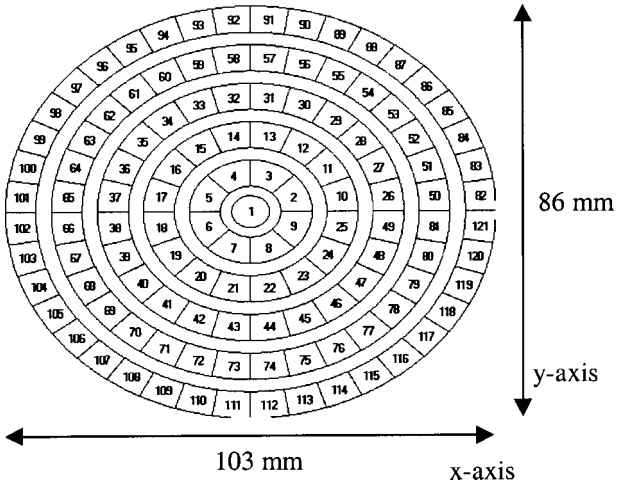


FIG. 1. Geometry of the transducer array.

III. THE TIME REVERSAL MIRROR AND THE DORT METHOD PRINCIPLE

The time reversal process and the DORT method which have been presented in several papers[8–18] are briefly reviewed in this section.

A. "Conventional" time reversal

A time reversal process allows the conversion of a divergent wave emerging from a source into a wave focusing on the source. In our case, the source is a flat bottom hole whose position is unknown. The time reversal process is an adaptive focusing technique that requires the following three steps:

(1) Transmission of a short widely divergent wave by several elements of the array; for this particular application, it was established that 25 central transducers was a good compromise.

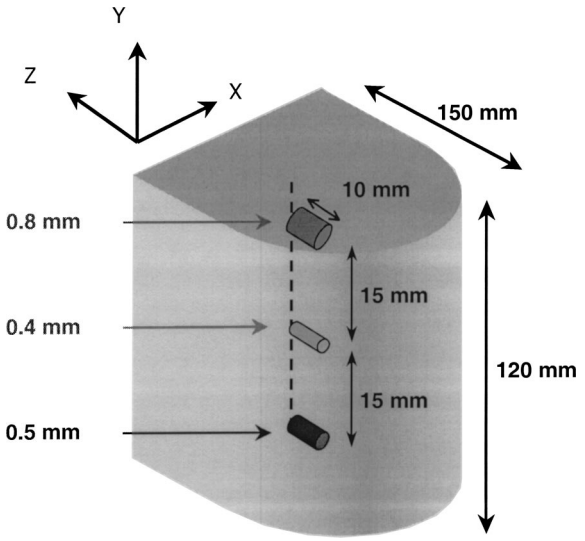


FIG. 2. The titanium sample.

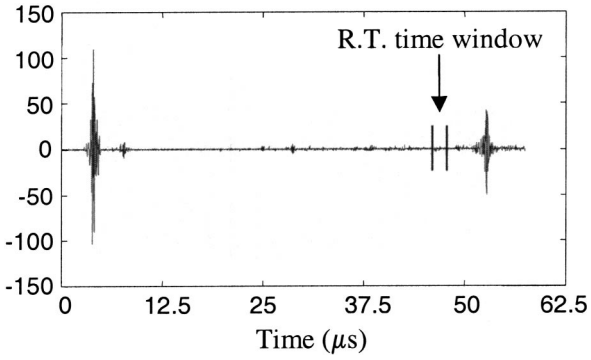


FIG. 3. Time reversal and DORT window.

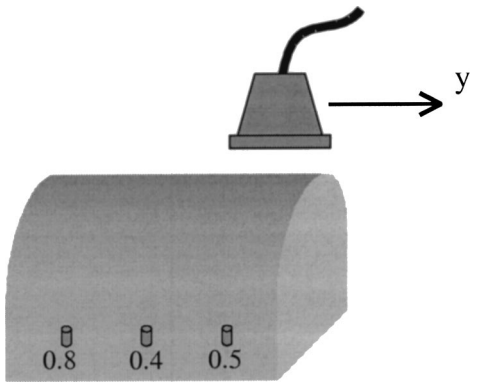


FIG. 4. Translation of the array along the *y* axis.

(2) The wave reflected by the defect is detected by the array and the echographic signals are recorded in the memories of the electronic system.
(3) These signals are then time reversed and used to transmit a wave focusing on the flaw. The time reversal process is performed on a 0.8 μs time window corresponding to the depth of the defect (Fig. 3).

In this paper, we study the temporal signals recorded by the TRM after two iterations. Let us describe two treatments applied to these signals. Because of 16 defective channels on the electronic system, only 105 signals are exploited. First, we present the incoherent summation $\mathrm{Inc}(t)$[11,12] of the 105 linear signals $[f_k(t)]_{1\leqslant k\leqslant 105}$, determined according to the summation

$$\mathrm{Inc}(t)=\sum_{k=1}^{105} f_k(t), \tag{1}$$

where $f_k(t)=R_{2,k}(t)$ is the ultrasonic signal recorded on the transducer k at the second iteration. This process is called ''incoherent summation'' [Eq. (1)] because the individual data are not in phase.

Another compact presentation of the time reversal process is the ''iterative coherent summation'' $\mathrm{Coh}(t)$ [Eq. (2)] of the linear signals $[f_k(t)]_{1\leqslant k\leqslant 105}$, determined according to the summation

$$\mathrm{Coh}(t)=\sum_{k=1}^{105} f_k(t-\tau_k), \tag{2}$$

where the delay τ_k is the time when the signal $R_{1,k}(t)$ recorded at the first iteration is the maximum. For a zone containing a defect, this process will greatly enhance the signal from the defect, even if it is not at the array focus. Examples of these two representations will be given in Sec. IV B.

B. The DORT method

The DORT method has been described in several papers.[13–15,17,18] This method is derived from the theoretical analysis of time reversal mirrors. It is based on a matrix formalism that describes a transmit–receive process performed with an array of L transducers. Such an array insonifying a given medium is a linear and time invariant input–output system characterized by $L\times L$ interelement impulse responses $k_{lm}(t)$ defined as the output of the element number l when the input of the element number m is a delta impulse. At a given frequency, the $L\times L$ transfer matrix $K(\omega)=[K_{lm}(\omega)]_{1<l,m<L}$ can be calculated by the Fourier transform of the responses $k_{lm}(t)$.

The time reversal operator, defined as $K^*(\omega)K(\omega)$, can be diagonalized. In general, the number of significant eigenvalues is equal to the number of well-resolved pointlike defects in the medium and each significant eigenvector of the time reversal operator is associated with one of the defects. The associated eigenvalue depends on the reflectivity of the defect and also on its position relative to the array. The first eigenvector generally corresponds to the strongest reflector and contains the signals that would be obtained after a large number of time reversal iterations.

In practice, the diagonalization of $K^*(\omega)K(\omega)$ is obtained from the calculation of the singular value decomposition (SVD) of the transfer matrix $K(\omega)$ [Eq. (3)],

$$K(\omega)=U(\omega)S(\omega)^tV^*(\omega), \tag{3}$$

where $S(\omega)$ is a real diagonal matrix of singular values. $U(\omega)$ and $V(\omega)$ are unitary matrices. The eigenvalues of $K^*(\omega)K(\omega)$ are the squares of the singular values of $K(\omega)$ and its eigenvectors are the columns of $V(\omega)$.

In summary, the DORT method consists of the measurement of the interelement impulse response functions $k_{lm}(t)$, the analysis of the singular values of the transfer matrix $K(\omega)$ and possibly the backpropagation of the eigenvectors associated with the most significant singular values.

IV. DETECTION IN A TITANIUM ALLOY SAMPLE

This part concerns the detection of the flat bottom holes in the titanium sample described above and illustrates the sensitivity of each technique to a bad positioning of the array.

For each technique, the input signal is a sine wave at 5 MHz modulated by a 1 μs Gaussian function. This signal is applied to one or several transducer(s) depending on the technique.

A first set of measurements is made to detect the three flaws. To this end, the array which is prefocused at the depth of the defects (Fig. 4), is translated parallel to the *y* axis in 1 mm steps.

A second set of measurements is made to test the robustness of each technique to a misalignment of the probe, that is

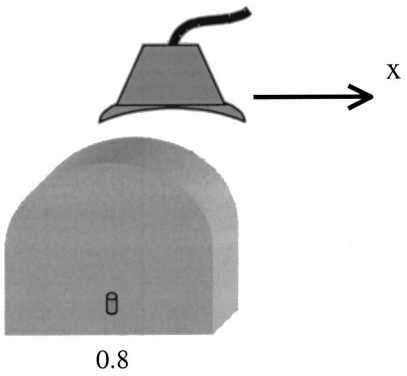


FIG. 5. Translation of the array along the x axis.

to say when the probe is no longer aligned with the liquid–solid interface. This is done for the 0.8 mm flaw by displacement of the array along the x axis (Fig. 5). The defect is initially at the array focus corresponding to $x=0$ mm. The data are acquired after translating the array along the x axis, by 0.1 mm steps and for $-3\ \text{mm} \leqslant x \leqslant 3\ \text{mm}$.

We first present the performance of the focusing technique in transmit–receive mode as if using a single large focused transducer.

A. Focusing in transmit–receive mode

The measurements are realized with the transducer array described in Sec. II but in order to simulate a single large focused transducer, the elements are fired simultaneously by the same input signal (focusing in transmit mode).

1. Detection of the three defects with the geometrical focusing (translation ∥ y)

For a given position of the array along the y axis, the received signals $[S_k(y,t)]_{1 \leqslant k \leqslant 105}$ are stored on each channel during 4 μs and summed to simulate a unique large transducer

$$A(y,t)=\sum_{k=1}^{105} S_k(y,t). \tag{4}$$

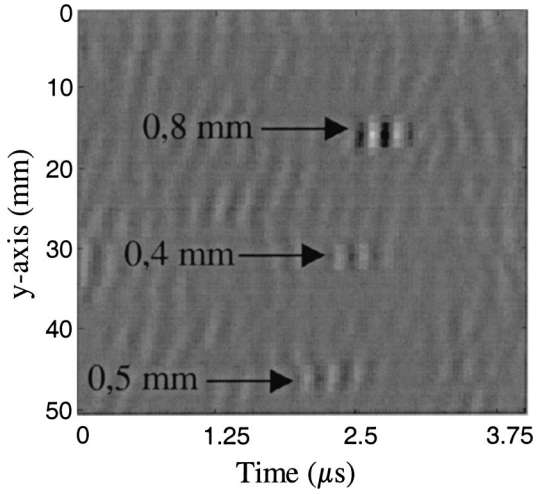


FIG. 6. Detection of the three defects with the focusing in transmit–receive mode.

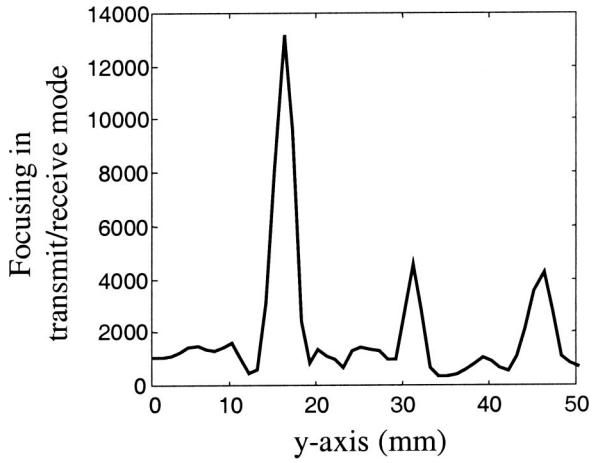


FIG. 7. Detection of the three defects with the focusing in transmit–receive mode at the depth of the defects on an 800 ns time window.

The function $A(y,t)$ is displayed in Fig. 6. The echoes from the 0.8 mm, 0.4 mm, and 0.5 mm defects are visible between 2 and 3 μs for the positions $y=16$, 31, and 46 mm, respectively. As can be seen on the figure, the array displacement is not perfectly parallel to the billet axis.

We then calculate the maximum of $A(y,t)$, defined as $M(y)$, for an 800 ns time window corresponding to the depth of the flaws and written as follows [Eq. (5)]:

$$M(y)=\max[A(y,t)]. \tag{5}$$

The function $M(y)$ is displayed at Fig. 7. The defects are well detected with a signal-to-noise ratio equal to 19, 10, and 9.5 dB for the 0.8, 0.5, and 0.4 mm flat bottom holes, respectively. The drawback of this focusing technique is that, due to the focused transmitted signal, a defect is detected by the transducer only when it is close to the focal point. Furthermore, as shown in the next paragraph, a misalignment of the transducer relative to the liquid–solid interface strongly degrades the focusing process.

2. Robustness of the geometrical focusing (translation ∥ x)

In this paragraph, we evaluate the detection quality of the 0.8 mm flaw with focusing in transmit–receive mode, after displacement of the array along the x axis (Fig. 5). The elements are simultaneously fired by the same input signal as before. The 105 signals are then acquired on an 800 ns time window corresponding to the defect position. This measurement is repeated after translating the array along the x axis. The association of a time window to a particular depth is possible because the array is prefocussed in the inspected area, which of course requires some knowledge on the sample's average acoustic properties.

We then calculate the $M(x)$ term [Eq. (5)], on an 800 ns time window corresponding to the depth of the flaw, for the 61 positions of the transducer (Fig. 8). The decrease of the detection level is equal to 11 dB after translating the array for a 0.5 mm distance. Note that the degradation observed in Fig. 8 is not perfectly symmetrical relative to the position $x=0$ mm, probably meaning a little asymmetry of the geometry.

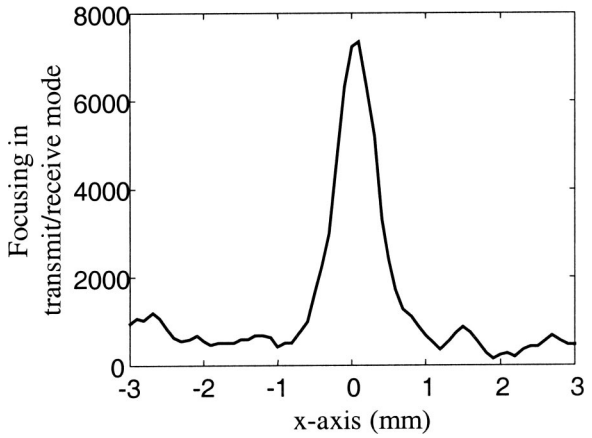


FIG. 8. Detection of the 0.8 mm defect with the focusing in transmit–receive mode along the *x* axis.

B. The conventional time reversal process

The time reversal mirror has been described in the Sec. III A. For the first transmission, the 25 central transducers are simultaneously fired by a short pulse. The time reversal process is performed on an 800 ns time window corresponding to the depth of the defects and the ultrasonic signals (B-Scan) are taken from the second iteration.

1. Detection of the three defects with the TRM (translation ∥ y)

Figure 9 displays the B-Scan measured at the second iteration when the 0.8 mm FBH is located 4 mm out of the array focus. The undulation of the wavefront reveals the out of focus position of the defect. The measurement of the B-Scan is done by translating the array along the *y* axis by 1 mm step (Fig. 4).

The Inc(y,t) and Coh(y,t) terms [Eqs. (1) and (2)] are calculated and we represent (Fig. 10) the peak of each summation for an 800 ns time window adapted to the depth of the defects. Three maxima are observed for $y=16$, 31, and 46 mm, associated with the flaws of 0.8, 0.4, and 0.5 mm diameter, respectively. The amplitude of each maximum is related to the size of the corresponding defect. The coherent summation generates three maxima wider and higher than these obtained with the incoherent summation. For example, by applying the coherent summation, the 0.8 mm FBH is detected at 15 mm distance from the focus. This is a great advantage of the self-adaptative property of the TRM process.

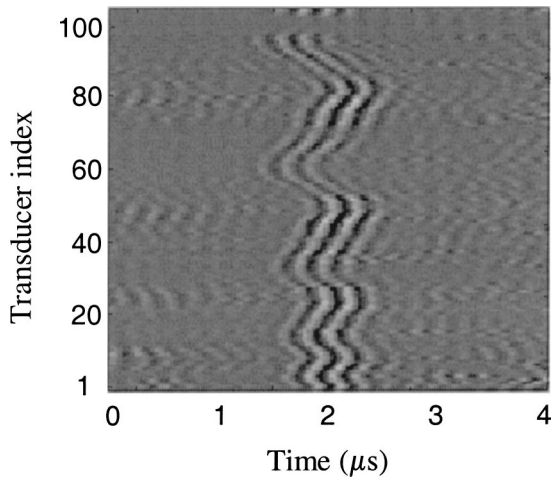


FIG. 9. Detection of the 0.8 mm defect located 4 mm from the array focus with the TRM.

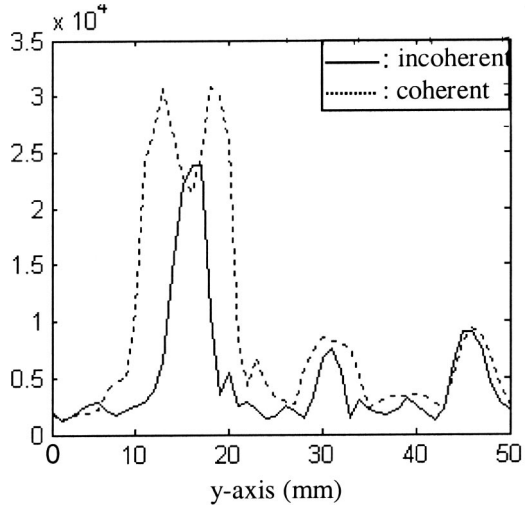


FIG. 10. Detection of the three defects with the incoherent and the coherent linear summations of the B-Scan measured with the TRM at the second iteration along the *y* axis.

2. Robustness of the TRM (translation ∥ x)

As for classical focusing, we evaluate the sensitivity of the time reversal process to a poor positioning of the transducer array relative to the titanium sample (Fig. 5). The peak in time of the Inc(x,t) and Coh(x,t) terms [Eqs. (1) and (2)] are calculated along the *x* axis (Fig. 11). They reach the maximum for $x=0$ mm, when the 0.8 mm FBH is at the focal point. Furthermore, the decrease of these incoherent and coherent summations is equal to 15 and 10 dB, respectively, after a displacement of the array for a 0.5 mm dis-

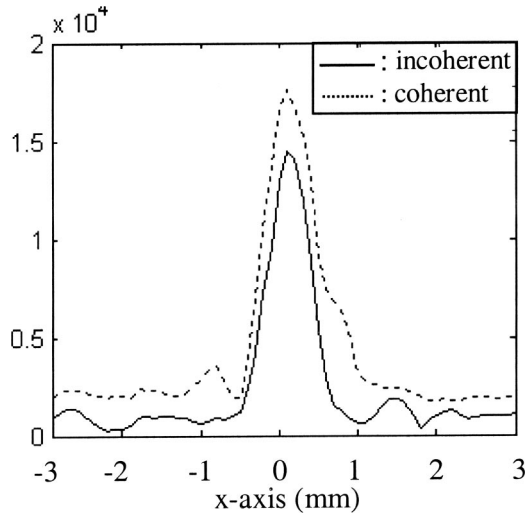


FIG. 11. Detection of the 0.8 mm defect with the incoherent and the coherent linear summations of the B-Scan measured with the TRM at the second iteration along the *x* axis.

tance. These compact representations emphasize the great advantage of the coherent summation for flaw detection out of the array focus.

These measurements were also made with nine transducers (the central transducer and the first crown) and one transducer (the central transducer) for the first transmission. We wondered whether the sensitivity of the time reversal process would be higher when the first transmission was less focused. In fact, for nine transducers, the results are similar to those obtained with 25 transducers whereas for one single transducer the energy transmitted is too weak and the defect never detected.

C. The DORT method

In order to evaluate the potentialities of the DORT method for flaw detection, it is applied to the titanium alloy sample in the same configuration as before (Figs. 1 and 2). The first step of the technique is the measurement of the interelement impulse responses $k_{lm}(t)$.

1. Detection of the three defects with DORT (translation ∥ y)

Figure 3 illustrates the interelement impulse response from element 10 to element 40 when the array is located in front of the 0.4 mm diameter FBH. The strong echoes coming from the front and the back faces of the titanium sample are easily visible. The transducer array is translated along the y axis by 1 mm steps (Fig. 4) and the interelement impulse responses $k_{lm}(t)$ are acquired on a wide temporal window. As the array is prefocused at 140 mm in the titanium sample, the choice of the temporal window corresponds to a slab at the same depth for each transducer.

2. Singular values at 5 MHz

The information corresponding to 140 mm depth inside the medium is selected and the transfer matrix $K(\omega)$ is calculated on an 800 ns time window. Working on such short signals is necessary to get a reasonable signal-to-noise ratio, otherwise the weak echo from the defect is drowned under the accumulation of the signal reflected by the grain structure.

The singular values are calculated at 5 MHz for each position of the array along the y axis (Fig. 12). The first singular value varies with the position of the flaws relative to the array focus. For example, the first singular value is associated with the FBH of diameter 0.8 mm, 0.4 mm, and 0.5 mm for the positions of the following array: $1\,\text{mm} \leqslant y \leqslant 25\,\text{mm}$, $26\,\text{mm} \leqslant y \leqslant 36\,\text{mm}$, $37\,\text{mm} \leqslant y \leqslant 49\,\text{mm}$, respectively. Furthermore, the differences between the three maxima reveal the different effective reflectivities of the flaws.

In particular, we can observe that the maximum associated with the 0.5 mm defect is smaller than the maximum associated with the 0.4 mm defect. This probably results from the fact that the 0.5 mm defect is not perfectly flat bottomed.

Near $y = 28\,\text{mm}$, two defects are simultaneously detected and a crossing can be observed on the two main singular values. For $y < 28\,\text{mm}$, the first and the second singular values are, respectively, associated with the 0.8 mm and the 0.4 mm defects while it is the contrary for $y \geqslant 28\,\text{mm}$.

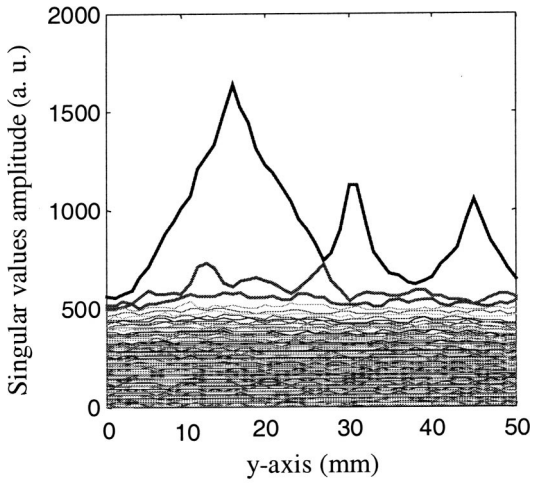


FIG. 12. Detection of the three defects with the singular values amplitude calculated at 5 MHz along the y axis.

The other singular values, called ''noise'' singular values, mostly result from the grain structure of titanium and electronic noise. Note that these ''noise'' singular values have no significant fluctuation along the y axis.

3. Frequency analysis

Until now, the study of the singular values was made at the central frequency of the transducers. However, the singular value decomposition may be done at any chosen frequency in the bandwidth of the transducer. Thus, it is interesting to observe the frequency dependence of the singular values. In our case, the SVD of the transfer matrix is calculated for the frequencies between 3.1 and 6.3 MHz at 0.16 MHz intervals.

For example, the singular values are represented when the 0.4 mm FBH is at the array focus and 5 mm out of focus, respectively (Figs. 13 and 14). For the first position, the first singular value is well separated from the others for frequencies above 4 MHz. The maximum is obtained at 5 MHz

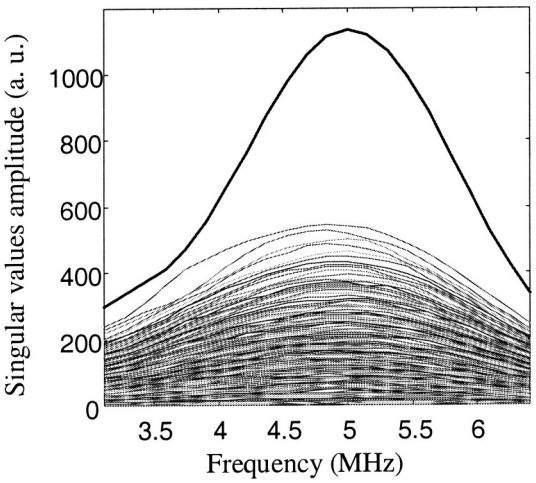


FIG. 13. Singular values amplitude when the 0.4 mm flaw is located at the array focus.

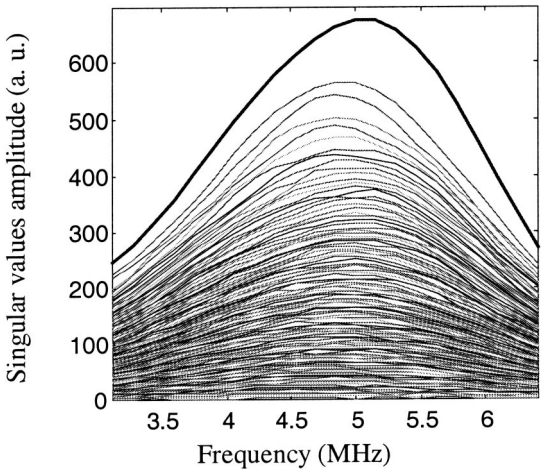


FIG. 14. Singular values amplitude when the 0.4 mm flaw is located 5 mm out of the array focus.

which is the central frequency of the transducers. When the defect is 5 mm out of focus, the first singular value associated with the defect, is closer to the other singular values.

For the second position of the defect, the first eigenvector $V_1(\omega)$ is calculated from 3.1 to 6.3 MHz with a frequency spacing equal to 0.16 MHz. Thus, the first ''temporal eigenvector'' $v_1(t)$ (Fig. 15) is obtained by inverse Fourier transform of $V_1(\omega)$. This ''temporal eigenvector'' represents the response from the defect to the array which is by reciprocity the set of signals that should be applied on each transducer in order to focus on the corresponding flaw. Although the 0.4 mm FBH is 5 mm away from the focal point, the temporal reconstruction of the first ''temporal eigenvector'' $v_1(t)$ remains associated with this defect. Indeed, the wavefront of this eigenvector shows undulating behavior which is explained by the out of focus position of the flaw.

4. Linear incoherent and coherent summations of the first temporal eigenvector $v_1(t)$

We have shown in Sec. III A a compact presentation of the time reversal process: the incoherent and coherent summations of the B-Scan acquired at the second iteration. The same process is also valid for the DORT method. Indeed, it consists of calculating the linear incoherent $\mathrm{Inc}(y,t)$ and coherent $\mathrm{Coh}(y,t)$ summations of the first ''temporal eigenvector'' $v_1(t)$ obtained for each position y of the array, according to Eqs. (1) and (2). In this case, $f_k(t)=a_{1,k}(y,t)$, where $a_{1,k}(y,t)$ is the temporal signal (A-Scan) extracted from $v_1(t)$ on the transducer k. The delay τ_k introduced in Eq. (2), corresponds to the time when the signal $a_{1,k}(y,t)$ is at the maximum.

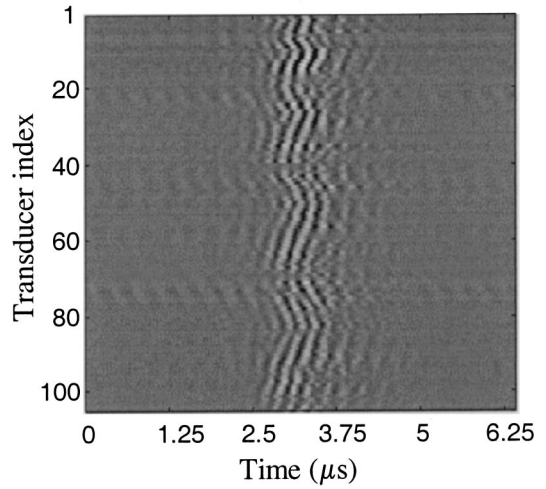


FIG. 15. First ''temporal eigenvector'' $v_1(t)$ when the 0.4 mm flaw is located 5 mm out of the array focus.

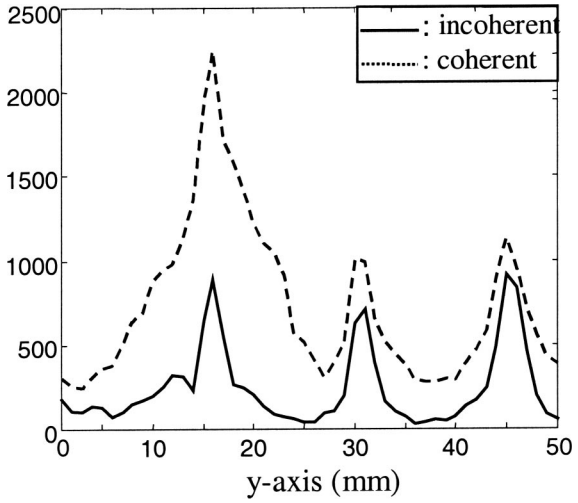


FIG. 16. Detection of the three defects with the incoherent and coherent summations of the first ''temporal eigenvector'' $v_1(t)$ calculated along the y axis.

We give an example of this process in Fig. 16 which presents the peak of the $\mathrm{Inc}(y,t)$ and $\mathrm{Coh}(y,t)$ terms, calculated for each position of the array along the y axis, by 1 mm spacing. We observe on each curve, as for the first singular value (Fig. 12), three maxima for the positions $y=16$, 31, and 45 mm associated with the 0.8, 0.4, and 0.5 mm flaws, respectively. As for time reversal, the incoherent summation produces three narrow peaks whereas the effect of the coherent summation is to widen the three lobes.

5. Robustness of the DORT method (translation ∥ x)

The performances of the focusing technique in transmit–receive mode and of the TRM have been investigated in Secs. IV A and IV B when the array, initially in front of the 0.8 mm flaw ($x=0$ mm) is translated along the x axis (Fig. 5). The DORT method is now applied to the same experimental procedure.

The interelement impulse responses $k_{lm}(t)$ are measured for the 61 positions of the transducer and stored for an 800 ns time window corresponding to the depth of the defect. The transfer matrix $K(\omega)$ is then calculated and decomposed in singular values for frequencies between 3.1 and 6.3 MHz with a frequency spacing equal to 0.16 MHz.

The singular values distribution is presented at 5 MHz in Fig. 17. The maximum of the first singular value associated with the defect is obtained for $x=0$ mm, and is still significantly higher than the noise level when the defect is far from the array focus. For example, the first singular value decreases by 5 dB for a displacement of the array equal to 1.5

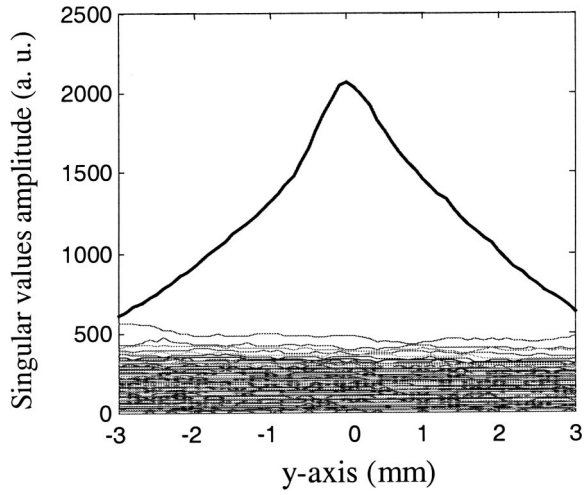


FIG. 17. Detection of the 0.8 mm flaw with the singular values amplitude at 5 MHz calculated along the x axis.

mm. Thus we see that the DORT method allows us to detect the flaw even if the transducer array is no longer adapted to the geometry of the sample.

The functions $\mathrm{Max}_t(\mathrm{Inc}(x,t))$ and $\mathrm{Max}_t(\mathrm{Coh}(x,t))$ present a maximum for $x=0.1$ and 0.2 mm, respectively (Fig. 18). Furthermore, these two terms lose 8 and 2.5 dB, respectively, for a displacement of the array equal to 0.5 mm. The coherent summation of $v_1(t)$ is a good way to break away from the loss of information when the array is no longer adapted to the geometry of the sample. Thus, it will be more interesting to analyze the coherent summation of $v_1(t)$ for a defect located out of the array axis.

This robustness is explained by the fact that for the DORT method, each transducer is fired separately, whereas the time reversal process simultaneously uses the 25 or 9 central transducers for the first transmission. Consequently, the transmission is less focused with the DORT method than with the time reversal process. Furthermore, the DORT technique takes advantage of a large angular diversity of incident

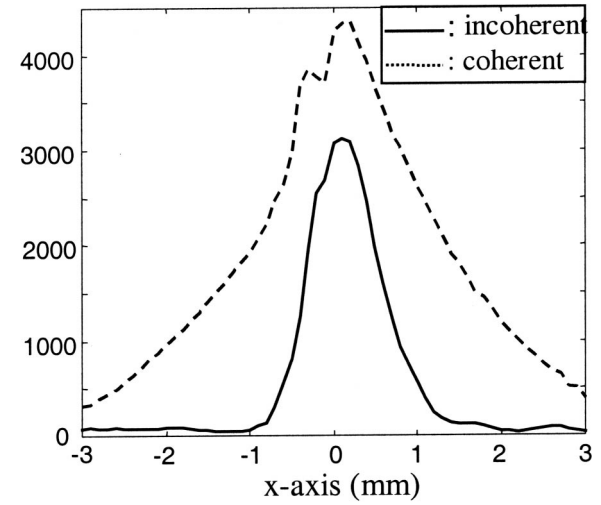


FIG. 18. Detection of the 0.8 mm flaw with the incoherent and coherent summations of the first ''temporal eigenvector'' $v_1(t)$ calculated along the x axis.

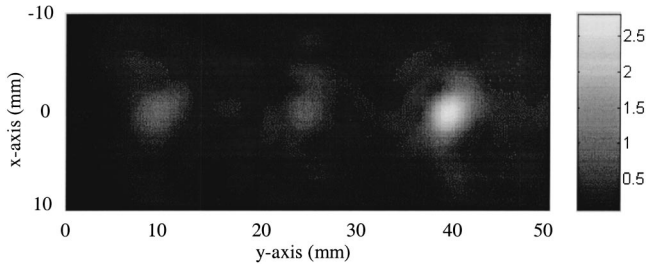


FIG. 19. Synthetic image of the sample at 140 mm depth with the time reversal mirror.

waves in the titanium sample, allowing a better detection than time reversal when the array is not perfectly aligned with the sample.

V. IMAGING IN THE TITANIUM SAMPLE

The procedures described in Sec. IV were only aimed at detecting the defects. Here, we want to make an image of the sample. To this end, we have to introduce some prior knowledge of the sample shape and we use a numerical backpropagation algorithms to reconstruct the sources of the echoes.

We propose to achieve defect localization using the simulation code PASS[18] (Phased Array Simulation Software) developed in our laboratory by Didier Cassereau. This software calculates the acoustic field produced by large arrays of transducers in complex media.

A. Imaging with the time reversal process

The time reversal process is exploited in order to produce a synthetic image of the sample, at the depth of the three flaws. The B-Scans measured along the y axis after two iterations (Fig. 9) are numerically backpropagated with the PASS code for the 50 positions of the array. Then, the 50 fields calculated along the x and y axis, are incoherently summed, taking into account the 1 mm array pitch between two consecutive measurements (Fig. 19). The image presents three maxima related to the reflectivity of each defect. Note that the 0.4 mm defect is detected with a 7.5 dB signal-to-noise ratio.

B. Imaging with the DORT method

The first ''temporal eigenvector'' obtained with DORT can be exploited as a B-Scan and thus numerically backpropagated. The first ''temporal eigenvector'' is backpropagated for the 50 positions of the array along the y axis, at a depth of 140 mm. The incoherent combination of the indi-

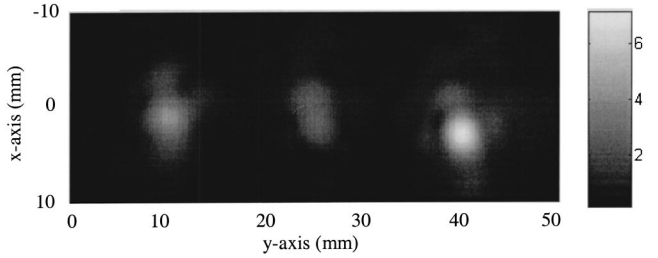


FIG. 20. Synthetic image of the sample at 140 mm depth obtained with the DORT method.

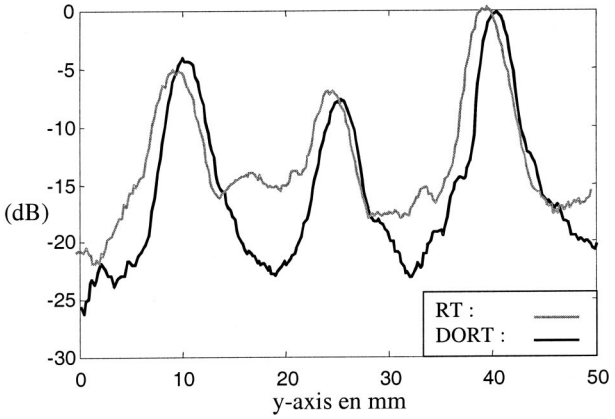


FIG. 21. Maximum of the projection of the two images along the y axis (dB).

vidual images calculated for each position of the array provides a complete image of the sample. Figure 20 is the synthetic image of the sample at the depth of the flaws. The three defects are detected with a good signal-to-noise ratio.

As shown in Fig. 21, the smallest flaw is detected 15 dB above the noise. The numerical backpropagation of the experimental data shows the efficiency of the DORT method in nondestructive testing to detect and locate a 0.4 mm diameter FBH in a noisy structure. This technique provides a synthetic image of the sample at a depth of 140 mm where the defects (including the 0.4 mm one) are well detected.

This study confirms the ability of the DORT method and of the time reversal mirror to detect small defects in a complex medium with a satisfactory signal-to-noise ratio.

VI. THE DORT METHOD IN A NOISE AREA

In Sec. IV C, we have applied the DORT method on 800 ns temporal window corresponding to the depth of the three defects. This detection technique is also applied to a defect free zone. The interelement impulse responses are acquired on a 4.2 μs temporal window, for 50 array positions ranging from 1 to 50 mm. For each position of the array, the transfer matrix is calculated on 18 temporal windows of 800 ns duration at 200 ns steps. This corresponds to depth ranging from 132 to 144 mm. After decomposition of the 50 times 18 matrices, the behavior of the singular values is analyzed.

The first singular value is represented at 5 MHz along the z and y axis, forming an image of the medium (Fig. 22). Three maxima associated with the defects of 0.8, 0.4, and 0.5 mm diameter appear around the depth $z=140$ mm for the positions $6\text{ mm}\leqslant y\leqslant 26\text{ mm}$, $28\text{ mm}\leqslant y\leqslant 33\text{ mm}$, and $41\text{ mm}\leqslant y\leqslant 49\text{ mm}$, respectively. Note that each flaw is detected for several temporal windows (5 for the 0.8 mm defect). This image also shows that a peak appears for several noise areas. In order to show the robustness of the DORT method, we now show that the peaks observed in noise areas cannot effectively correspond to any defect.

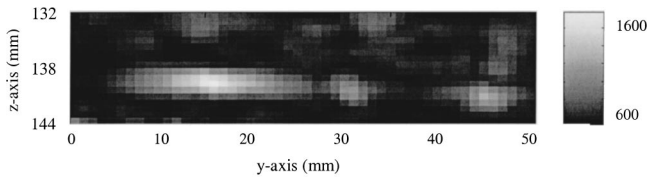


FIG. 22. Image of the medium with the first singular value calculated versus z and y axis.

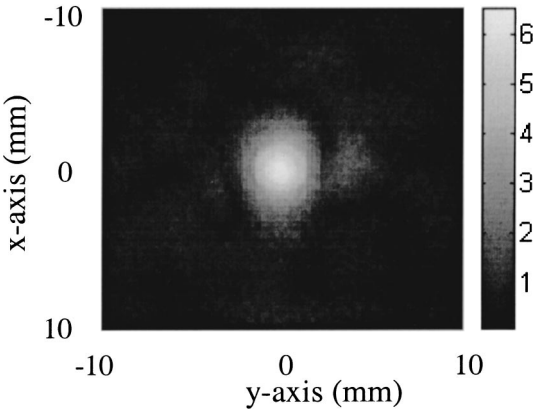


FIG. 23. Acoustic field versus x and y axis, obtained by backpropagation of the first "temporal eigenvector" at the depth of $z=132$ mm for the array position $y=33$ mm.

We numerically backpropagate the first eigenvector $v_1(t)$ corresponding to a noise area for several positions of the array. For instance, we choose to backpropagate $v_1(t)$ calculated for four array positions along the y axis ($y=32$, 33, 34, and 35 mm) and at the depth of $z=132$ mm (e.g., corresponding to the first temporal window). The acoustic field obtained by backpropagation of $v_1(t)$ for $y=33$ and 34 mm is represented along the x and y axis (Figs. 23 and 24). Note that the geometrical focal point is at the center of the image (e.g., for $x=y=0$ mm) for each backpropagation. For each array position ($32\text{ mm}\leqslant y\leqslant 35\text{ mm}$), the field calculated along $x=0$ axis presents a maximum at the focal point (Fig. 25).

Thus, the behavior of the DORT method is completely different in a high speckle zone than the zone around a defect. For the last case, the field is focused on the defect position. Consequently, when the array is translated for a 1 mm distance along the y axis, the distance between the two consecutive peaks is also equal to 1 mm. Moreover, the maximum associated with a defect coincides with the focal

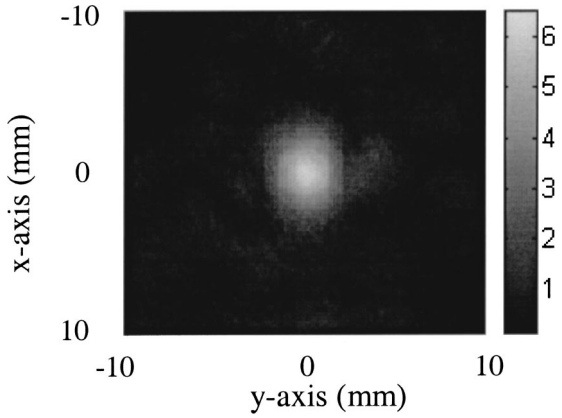


FIG. 24. Acoustic field versus x and y axis, obtained by backpropagation of the first "temporal eigenvector" at the depth of $z=132$ mm for the array position $y=34$ mm.

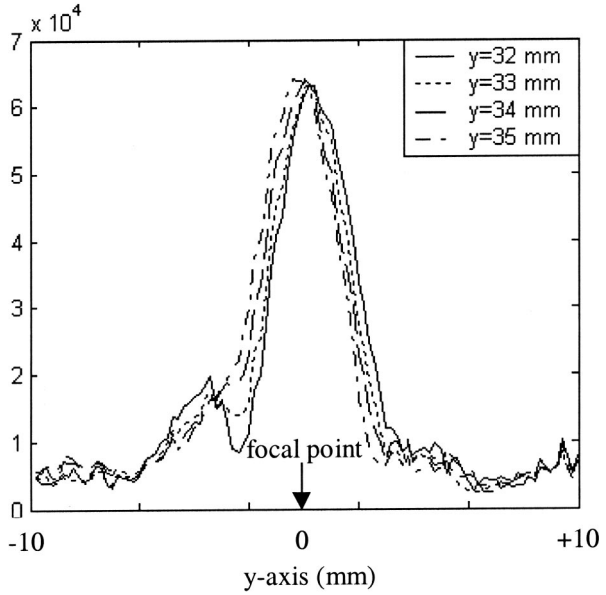


FIG. 25. Acoustic field along $x=0$ axis, obtained by backpropagation of the first ''temporal eigenvector'' at the depth of $z=132$ mm and for the array positions $y=31$, 32, 33, 34, 35 mm.

point only when the defect is at this point. Concerning the noise area, we have shown that the first eigenvector always focuses at the array focal point, because the array used for these experiments is prefocussed. This analysis allows us to discriminate the contribution of a defect from the speckle caused by the microstructure.

VII. CONCLUSION

Three techniques have been applied to the detection of FBH in a titanium billet, the standard transmit/receive focusing, the time reversal mirror and the DORT method. The DORT technique provides a real improvement in detection of deep weak defects in a noisy structure. On the synthetic images obtained by numerical backpropagation of the data with PASS, a 0.4 mm large and 140 mm deep defect was detected with a 15 dB signal-to-noise ratio using the DORT method whereas it was 7.5 dB for the time reversal process.

The sensitivity of each detection technique to a misalignment of the array relative to the liquid–solid interface has also been studied. The standard focusing technique does not allow a correct detection of the 0.8 mm flaw when the array is no longer adapted to the interface geometry. The sensitivity of the time reversal mirror is better than the standard focusing technique but worse than the DORT method. Thus, the DORT technique promises to be useful to detect flaws located in samples with complex geometry and using arrays which are not necessarily matched to this geometry.

We have also shown that the numerical backpropagation of the first eigenvector is a good test to get rid of false alarms. In a noise area presenting a peak on the first singular value, it appears that the backpropagation of the corresponding ''noise eigenvectors'' produces a peak which remains at the array focus while moving the array.

These detection techniques present several advantages and disadvantages. The DORT method and the standard transmit/receive focusing technique do not require programmable generators which are necessary for the time reversal process. The measurement of the interelement impulse responses for the DORT method provides information on the whole depth of the sample, whereas the iterative TR process requires an iteration for each short time windows. On the other hand, the DORT method requires N transmissions for N transducers, while the time reversal process only requires three transmissions. Therefore, further studies should focus on minimizing the data volume, measurements and calculations times.

This work does not answer the difficult question ''to what degree of inhomogeneities will the method remain efficient?'' Although some answer was given by Borcea *et al.*[19] in a theoretical and numerical analysis of a 2D fluid media, we believe that an answer to this question in the context of NDE should be the topic of further studies.

[1] A. Li, R. Roberts, F. J. Margetan, and R. B. Thompson, ''Study of the effect of microstructure on ultrasonic signal attenuation,'' in *Reviews of Progress in QNDE*, Vol. 20, edited by D. O. Thompson and D. E. Chimenti (Melville, New York, 2001), pp. 1322–1329.

[2] S. Chatillon, G. Cattiaux, M. Serre, and O. Roy, ''An ultrasonic system for the inspection of components with irregular geometry using a flexible phased array contact transducer and related processing algorithms,'' in *Reviews of Progress in QNDE*, Vol. 19, edited by D. O. Thompson and D. E. Chimenti (Melville, New York, 2000), pp. 1095–1102.

[3] P. J. Howard, D. C. Copley, J. D. Young, E. J. Nieters, and R. S. Gilmore, ''An improved methodology for the inspection of titanium alloys,'' IEEE Ultrason. Symp. Proc. **1**, 727–730 (1996).

[4] V. Lupien and F. Cancre, ''Ultrasonic phased array inspection of titanium billets,'' in *Reviews of Progress in QNDE*, Vol. 20, edited by D. O. Thompson and D. E. Chimenti (Melville, New York, 2001), pp. 919–926.

[5] C. W. Sherwin, J. P. Ruina, and R. D. Rawcliffe, ''Some early developments in synthetic aperture radar systems,'' IEEE Trans. Mil. Electron. **6**, 111–115 (1962).

[6] M. Karaman, P.-C. Li, and M. O'Donnell, ''Synthetic aperture imaging for small scale systems,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 429–442 (1995).

[7] F. J. Margetan, H. Wasan, and R. B. Thompson, ''An experimental study of microstructure-induced ultrasonic signal fluctuations in jet-engine titanium alloys,'' in *Reviews of Progress in QNDE*, Vol. 19, edited by D. O. Thompson and D. E. Chimenti (Melville, New York, 2000), pp. 1433–1439.

[8] M. Fink, ''Time reversal of ultrasonic fields—Part I: Basic principles,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control, **39**, 555–566 (1992).

[9] F. Wu, J. L. Thomas, and M. Fink, ''Time reversal of ultrasonic fields-Part II: Experimental results,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 567–578 (1992).

[10] M. Fink, C. Prada, and F. Wu, ''Self focusing in inhomogeneous media with time reversal acoustic mirrors,'' in *Proceedings of the IEEE Ultrasonic Symposium, 1989*, edited by B. R. McAvoy, 1989, Vol. 2, pp. 681–686.

[11] N. Chakroun, M. Fink, and F. Wu, ''Time reversal processing in ultrasonic nondestructive testing,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 1087–1098 (1995).

[12] V. Miette, L. Sandrin, F. Wu, and M. Fink, ''Optimisation of time reversal processing in titanium inspections,'' IEEE Ultrason. Symp. Proc. **1996**, 643–647 (1996).

[13] C. Prada, J.-L. Thomas, and M. Fink, ''The iterative time reversal process: Analysis of the convergence,'' J. Acoust. Soc. Am. **97**, 62–71 (1995).

[14] N. Mordant, C. Prada, and M. Fink, ''Highly resolved detection and selective focusing in a waveguide using the D.O.R.T. method,'' J. Acoust. Soc. Am. **105**, 2634–2642 (1999).

[15] C. Prada, S. Manneville, D. Spoliansky, and M. Fink, ''Decomposition of the time reversal operator: Detection and selective focusing on two scatterers,'' J. Acoust. Soc. Am. **99**, 2067–2076 (1996).

[16] E. Kerbrat, R. K. Ing, C. Prada, D. Cassereau, and M. Fink, ''The D.O.R.T method applied to detection and imaging in plates using Lamb waves,'' in

Review of Progress in QNDE, edited by D. O. Thompson and D. E. Chimenti (Melville, New York, 2000), pp. 934–940.
[17] C. Prada and M. Fink, ''Separation of interfering acoustic scattered signals using the invariant of the time-reversal operator. Application to Lamb waves characterization,'' J. Acoust. Soc. Am. **104**, 801–807 (1998).
[18] D. Cassereau, ''Le retournement temporel en acoustique—Théorie et modélisation,'' thèse d'habilitation, Paris 7, 1997; see also http://www.loa.espci.fr/pass
[19] L. Borcea, G. Papanicolaou, C. Tsogka, and J. Berryman, ''Imaging and time reversal in random media,'' Inverse Probl. **18**, 1247–1279 (2002).

Analysis of flexural mode focusing by a semianalytical finite element method

Takahiro Hayashi[a] and Koichiro Kawashima
Department of Mechanical Engineering, Nagoya Institute of Technology, Gokiso Showa Nagoya 466-8555, Japan

Zongqi Sun and Joseph L. Rose[b]
Department of Engineering Science and Mechanics, The Pennsylvania State University, 411E Earth & Engineering Science Building, University Park, Pennsylvania 16802

(Received 24 May 2002; revised 13 December 2002; accepted 16 December 2002)

Focusing is one of the most promising techniques for the detection of small defects in pipe works, in which guided waves including longitudinal and flexural modes are tuned so that ultrasonic energy can be focused at a target point in a pipe, and analysis of reflected waves gives information on defects such as location and size. In this paper, the focusing technique is discussed by way of a simulation of guided wave propagation in pipe by a semianalytical finite element method (SAFE). Experiments and SAFE calculations were compared for waveforms transmitted by a single transducer and received at different circumferential positions initially, and then the focusing phenomena were experimentally observed using focusing parameters obtained by calculations. Calculation and visualization were conducted to clarify focusing phenomena in pipe in investigating the potential of the focusing technique. These results show that the time-reversal idea helps in understanding focusing and that resolution of focusing is strongly affected by incident waveforms as well as the number of channels available in an experiment. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1543931]

PACS numbers: 43.20.Gp, 43.20.Px [LLT]

I. INTRODUCTION

Since thousands of miles of tubing and piping are used in all sorts of power generation, petrochemical, and manufacturing plants, nondestructive evaluation has become an important subject for reduced inspection time and costs. Ultrasonic nondestructive evaluation by guided waves propagating over long distances in a pipe has recently attracted considerable attention as a useful means for solving these issues; many theoretical and experimental studies have been published.[1–11]

All of these works on guided waves in a pipe are based on a theoretical study of the normal mode expansion presented by Gazis in 1959,[12] in which dispersion curves for a hollow cylinder with outer- and inner traction-free surfaces were derived and fundamental useful knowledge was presented on guided waves in a pipe. In order to achieve our aim of practical applications to quantitative evaluation for cracks in a pipe, however, more detailed theoretical or numerical studies have been required. The focusing technique, especially, proposed by Li and Rose[3,4] is expected to represent a key technique for defect detection in a pipe. The approach, however, is very hard to understand intuitively from the initial fundamental theoretical studies.

The basic idea of the focusing technique is as follows. First, a displacement distribution in the circumferential direction (circumferential profile) is predicted theoretically for partial loading by a single transducer. Circumferential profiles for multiple transducers can then be expressed by a superposition of this single transducer profile. Then, by tuning amplitudes and time delays of each channel independently, waveforms are enhanced at some predetermined target in a pipe. Since ultrasonic energy is converged on the target, a large reflected echo can be detected only when a defect is located at the target. Scanning various targets can take place by changing amplitudes and time delays. Observation of reflected waves then provides useful information on defects in a pipe, such as location and size.

The purpose of this study is to analyze this focusing technique by way of simulation and visualization of guided wave propagation.

Numerical calculation modeling such as finite element method (FEM) and boundary element method (BEM) is suitable for simulating ultrasonic wave propagation and to investigate it in detail. Moreover, visualizing these calculation results helps our understandings in such complicated wave propagation as guided waves. The semianalytical finite element method (SAFE), in which wave propagation in the longitudinal and circumferential direction is expressed by a superposition of orthogonal harmonics, is one of the most effective modeling techniques for carrying out large computations for guided wave propagation in a pipe. This special modeling technique designed for bar- and plate-like objects was developed by Cheng[13] and Zienkiewicz.[14] Recently, Datta[15–18] and Dong[19,20] have made significant progress on guided waves using SAFE. The authors[21] have studied complicated Lamb waves propagating in laminated plates using calculation and visualization techniques.

In this paper, the SAFE is used for simulation of cylin-

[a]Electronic mail: hayashi@megw.mech.nitech.ac.jp
[b]Electronic mail: jlresm@engr.psu.edu

drical guided wave propagation. First, calculated waveforms are compared with experiments to confirm the accuracy of the calculations and SAFE calculations are applied to focusing experiments. Next, we investigate the focusing status and its application to defect detection from visualization results. Moreover, the potential values of the focusing technique are discussed.

II. PRELIMINARY

A. Semianalytical FEM

In the SAFE used here, waveforms in circumferential (θ) and longitudinal (z) directions represent orthogonal harmonic functions, and the analytical dimensions are then reduced into one dimension in the thickness direction (r). Therefore, even guided wave propagation over long distance can be calculated with short calculation times and a small computational memory. Since the basic formulations are written in the references[14–18] in detail, only some fundamental and additive equations will be represented in this section. Displacement vector $\mathbf{U}$, containing displacements at nodes divided in the r direction, is represented using orthogonal harmonic functions $\exp(in\theta)$, $\exp(i\xi z)$, and $\exp(-i\omega t)$ in the r, θ, and t (time) direction, respectively

$$\mathbf{U}(\theta,z,t)=\int_{-\infty}^{+\infty}\overline{\mathbf{U}}(\theta,z)\exp(-i\omega t)d\omega, \tag{1}$$

$$\overline{\mathbf{U}}(\theta,z)=\sum_{n=-\infty}^{+\infty}\overline{\mathbf{U}}_n(z)\exp(in\theta), \tag{2}$$

$$\overline{\mathbf{U}}_n(z)=\int_{-\infty}^{+\infty}\tilde{\overline{\mathbf{U}}}_n\exp(i\xi_n z)d\xi, \tag{3}$$

where $\overline{\mathbf{U}}$ denotes the Fourier transform of $\mathbf{U}$ in terms of time t, and means the complex amplitude for harmonic waves at the angular frequency ω. $\overline{\mathbf{U}}_n$ is a complex Fourier series of $\overline{\mathbf{U}}$ in terms of circumferential direction θ, $\tilde{\overline{\mathbf{U}}}_n$ is the Fourier transform of $\overline{\mathbf{U}}_n$ in terms of longitudinal direction z, and ξ_n is the wave number propagating into the $\pm z$ directions for the nth circumferential function.

Similar to ordinary FEM, the virtual work principle or Hamilton's principle leads to the governing equations. The insertion of Eq. (1) in the governing equations gives eigensystems for the nth circumferential harmonic. Solving this eigensystem yields $2M$ eigenvalues ξ_{nm} and $2M$ right eigenvectors Φ^R_{nm} ($2M$ vector), where M is the dimension of the nodal displacement vector U and three times the number of nodes, and $m=1,2,...,2M$. The eigenvalue ξ_{nm} corresponds to the wave number of the mth mode and nth family. When it is real, it means a propagating mode, and an evanescence mode is represented for complex ξ_{nm}. Calculating the eigenvalues ξ_{nm} at many frequency steps yields dispersion curves for the nth family.

Here, following Ref. 19, the displacement vector $\overline{\mathbf{U}}_n$ for the nth circumferential harmonic (nth family) is obtained as

$$\overline{\mathbf{U}}_n=\sum_{m=1}^{M}\alpha_{nm}\Phi^{R\ \text{upper}}_{nm}\exp(i\xi_{nm}z), \tag{4}$$

where α_{nm} is an amplitude of the (n,m) mode to be obtained from boundary conditions, and $\Phi^{R\ \text{upper}}_{nm}$ is the vector consisting of upper M elements of Φ^R_{nm}, representing displacement distribution in the r direction. Equation (4) exhibits that $\overline{\mathbf{U}}_n$ is described as a superposition of guided wave modes for the nth circumferential harmonic.

Now suppose that a line source is located at $z=z_s$, $|\theta|<\theta_0$ [Fig. 1(a)] and that the displacement vector is described as

$$\overline{\mathbf{U}}^{\text{lineS}}_n=\sum_{m=1}^{M}\alpha^{\text{lineS}}_{nm}\Phi^{R\ \text{upper}}_{nm}\exp\{i\xi_{nm}(z-z_s)\}. \tag{5}$$

This solution has been obtained in Ref. 19. Then, the solution for the angle wedge transducer of incident angle θ_i and located at $|\theta|<\theta_0$, $|z-z_s|<z_0$ can be obtained as an integration of Eq. (5). See Fig. 1(b). By considering the phase shift due to incident angle, only amplitude α_{nm} should be replaced as

$$\alpha_{nm}\rightarrow\alpha_{nm}\times\frac{2\sin\{(\xi_0-\xi_{nm})z_0\}}{\xi_0-\xi_{nm}}, \tag{6}$$

where ξ_0 is represented as $\xi_0=\xi_a\sin\theta_i$ when assuming the wave number in the wedge is ξ_a.

When the angle wedge receiver is the same as the transmitter with the width $2\theta_0$ and $2z_0$ in the θ and z directions, respectively, the amplitude of the received signal is obtained as

$$\alpha_{nm}\rightarrow\alpha_{nm}\times\frac{2\sin\{(\xi_{nm}-\xi_0)z_0\}}{\xi_{nm}-\xi_0}\times\frac{2\sin n\theta_0}{n}. \tag{7}$$

When the excitation point is $\theta=\theta_s$ instead of $\theta=0$

$$\alpha_{nm}\rightarrow\alpha_{nm}\times\exp(-in\theta_s) \tag{8}$$

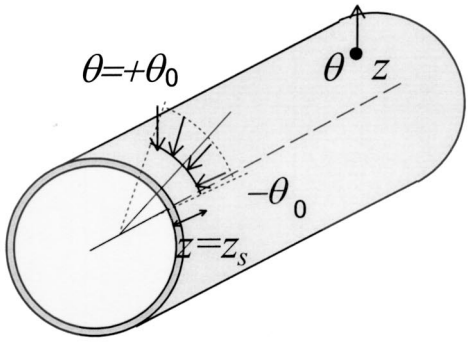
gives the solutions.

To summarize, when the angle wedge transmitter with incident angle θ_i and width of $2z_0$ and $2\theta_0$ are installed at $z=z_s$, $\theta=\theta_s$, displacements at the angle wedge receiver with receiving angle θ_i and width of $2z_0$ and $2\theta_0$ [Fig. 1(c)] are given as

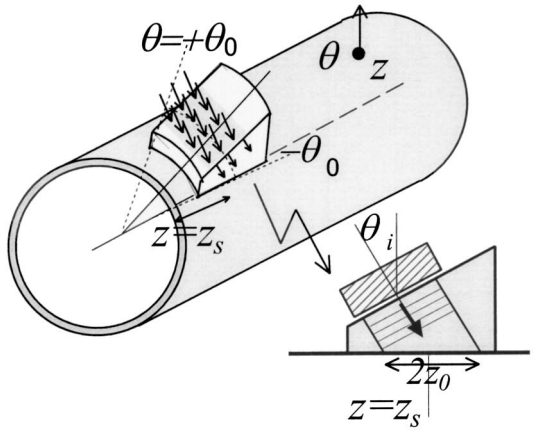
$$\begin{aligned}\overline{\mathbf{U}}^{\text{angle}S+\text{angle}R}_n&=\sum_{m=1}^{M}\alpha^{\text{angle}S+\text{angle}R}_{nm}\Phi^{R\ \text{upper}}_{nm}\\&\quad\times\exp\{i\xi_{nm}(z-z_s)\},\\ \alpha^{\text{angle}S+\text{angle}R}&=\alpha^{\text{line}S}\times\frac{2\sin\{(\xi_0-\xi_{nm})z_0\}}{\xi_0-\xi_{nm}}\\&\quad\times\frac{2\sin\{(\xi_{nm}-\xi_0)z_0\}}{\xi_{nm}-\xi_0}\times\frac{2\sin n\theta_0}{n}\\&\quad\times\exp(-in\theta_s).\end{aligned} \tag{9}$$

B. Focusing algorithm

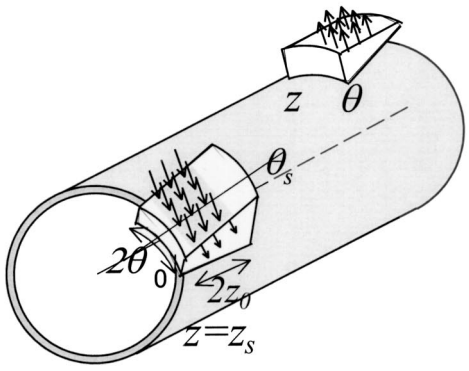
N transmitters and N receivers are located in the circumferential direction and numbered 1,2,...,N from the top of the pipe as shown in Fig. 2. Then, the complex amplitude of a harmonic wave at the frequency ω traveling from the ith transmitter to the jth receiver is defined as H_{ij}. H_{ij} is pre-



(a) Line source and point receiver.
Displacement field is described by Eq.(5).



(b) Angle wedge transmitter and point receiver.
Displacement field is described in Eq.(6).



(c) Angle wedge transmitter and angle wedge receiver.
Displacement field is described by Eq. (9)

FIG. 1. Boundary conditions.

determined by experiments or calculations, containing the information of amplitude and phase. Controlling the output level (amplitude) A_i and time delay t_i of the ith transmitter, the received signal is described as $a_i H_{ij}$, where

$$a_i = A_i \exp(-i\omega t_i). \tag{10}$$

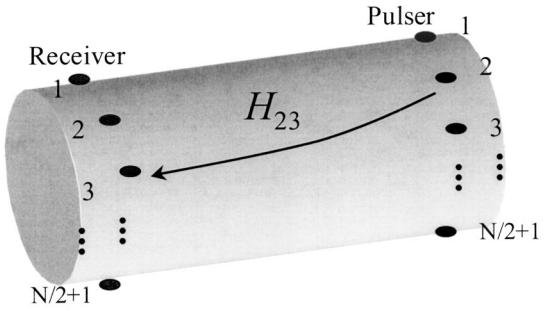


FIG. 2. Focusing algorithm.

Tuning all amplitudes and time delays from 1 to N, signals at the jth receiver q_j can be described as a superposition of $a_i H_{ij}$ as

$$q_j = \sum_{i=1}^{N} a_i H_{ij}. \tag{11}$$

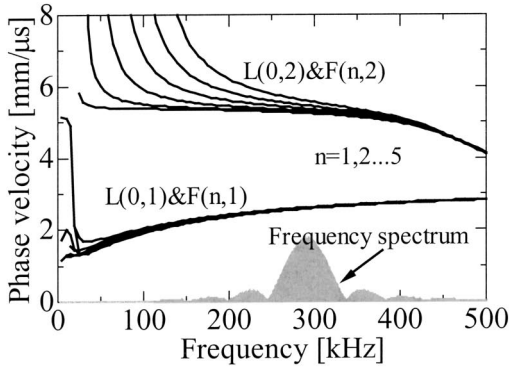
This is represented in the matrix form as

$$\mathbf{q} = \mathbf{Ha}, \tag{12}$$

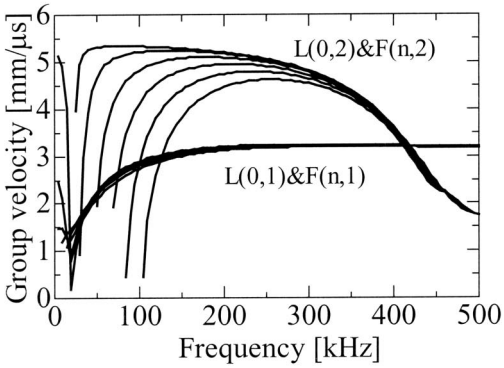
$$\mathbf{q} = [q_1 \ q_2 \ q_3 \ \cdots \ q_N]^T, \quad \mathbf{a} = [a_1 \ a_2 \ a_3 \ \cdots \ a_N]^T.$$

Here, letting the received signals be

$$\mathbf{q} = [1 \ 0 \ 0 \ \cdots \ 0]^T, \tag{13}$$

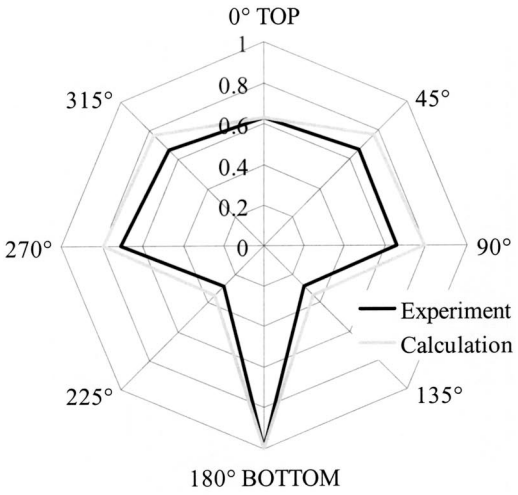


(a) Phase velocity

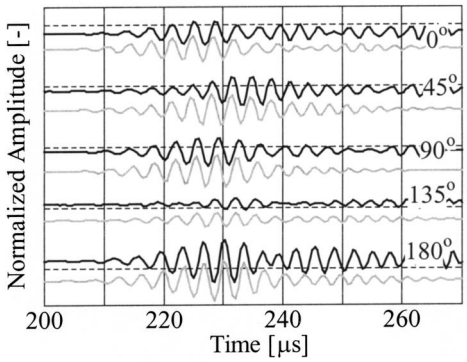


(b) Group velocity

FIG. 3. Dispersion curves for a steel pipe used in this paper; 88-mm outer diameter and 5-mm thickness. (Gray region denotes the frequency spectrum of incident wave, 290 kHz, 7.5-cycle tone-burst waves.)

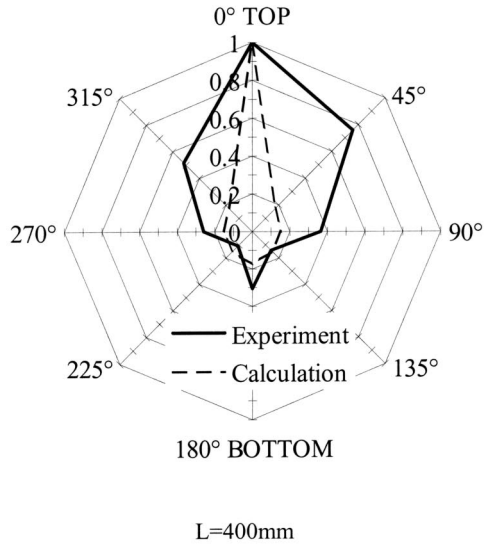


(a) Circumferential profile

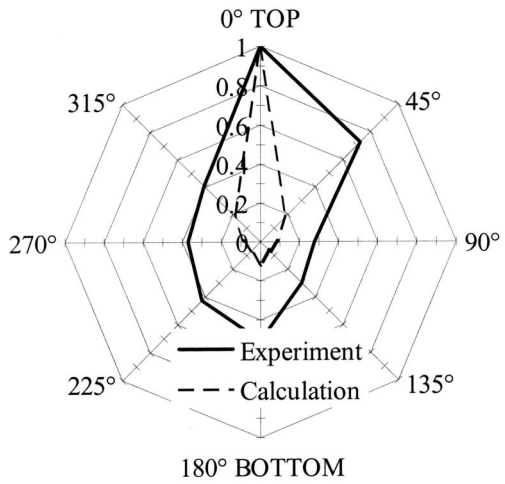


(b) Time domain profiles

FIG. 4. Experimental (black line) and calculated (gray line) waveforms at different circumferential positions at L=800 mm.



L=400mm



L=800mm

FIG. 5. Circumferential profile at focal points of L=400 mm and L=800 mm.

then we have a complex amplitude

$$\mathbf{a}=\mathbf{H}^{-1}\mathbf{q}. \tag{14}$$

Complex amplitude components a_i are represented by amplitudes and time delays as shown in Eq. (10).

From the predefined target amplitude **q** and the matrix **H** determined by numerical calculations, necessary amplitudes and time delays for each element are calculated. When a pipe in question is geometrically symmetric like a straight homogeneous pipe, matrix **H** becomes a cyclic matrix as

$$\begin{aligned} &H_{11}=H_{22}=H_{33}=\cdots=H_{NN},\\ &H_{12}=H_{23}=H_{34}=\cdots=H_{N1},\\ &\cdots, \end{aligned} \tag{15}$$

and then the deriving Eq. (14) is equivalent to the convolution algorithm developed by Li.[4]

III. COMPARISON OF SAFE CALCULATION AND EXPERIMENT

The following two experiments were carried out in this section in order to verify the SAFE calculations. The first one was a comparison between experimental and calculated flexural modes that were exited by a single transducer and received at various circumferential positions. The second one was a confirmation of the focusing phenomena by using amplitudes and time delays obtained from the SAFE calculations of circumferential profile H_{ij} and the focusing algorithm.

In these experiments, a steel pipe of 88-mm diameter and 5-mm thickness is used. Eight angle wedge transmitters with an incident angle of 30° are fixed at the end of the pipe and an angle wedge receiver with 30° receiving angle can be moved to receive waveforms at different circumferential and longitudinal positions.

Figure 3 shows dispersion curves of the steel pipe used in these experiments as well as a frequency spectrum of incident signals of about 290 kHz center frequency. L and F in Fig. 3 denote Longitudinal and Flexural modes, respectively. The numbers in parentheses correspond to (n,m) in Eq. (4), representing the mth mode of the nth family. The incident and receiving angles of 30° are suitable for exciting and

TABLE I. Amplitude and time-delay computation for focusing at the top for L=400 mm and 800 mm.

L=400 mm Top	Top, 0°	45°	90°	135°	Bottom, 180°	225°	270°	315°
Amplitude [-]	1.00	0.98	0.92	0.87	0.83	0.87	0.92	0.98
Time delay [μs]	2.38	2.04	1.22	0.00	1.88	0.00	1.22	2.04
L=800 mm Top	Top, 0°	45°	90°	135°	Bottom, 180°	225°	270°	315°
Amplitude [-]	0.48	0.60	0.72	0.35	1.00	0.35	0.72	0.60
Time delay [μs]	1.78	2.52	1.39	1.50	0.00	1.50	1.39	2.52

detecting L(0,2) and F(n,2) modes. In the first single transmitter tests, waveforms at eight circumferential positions (0°,45°,90°,135°,180°,225°,270°,315°) were detected in through-transmission configurations with a fixed angled transmitter on the top (0°) of the pipe and a receiving transducer at different positions. In the second focusing tests, a Matec multichannel system independently tuned amplitudes and time delays for eight angle wedge transmitters located at regular intervals in the circumferential direction. Circumferential profiles were measured by scanning the receiving transducer in the circumferential direction.

In the SAFE calculations, solving the eigensystems determined by given material properties and geometries yields wave numbers ξ_{nm} and wave structure vectors $\Phi_{nm}^{R\ \mathrm{upper}}$ in Eq. (4), and the boundary conditions give amplitudes α_{nm}. Then, displacements in Eq. (4) were given for each circumferential harmonic n and frequency step ω. Waveforms at any points $\mathbf{U}(\theta,z,t)$ are obtained by substituting displacements Eq. (4) into Eq. (2) and then into Eq. (1). Tone-burst signals of 290 kHz and 7.5 cycles were adopted as an excitation stress on the surface of the pipe so as to meet experimental conditions. So, boundary conditions in these calculations are traction-free on both inner and outer surface of the pipe and tone-burst normal stress applied on the loading region. The other geometric parameters such as transducer width (z_0 $=7.5\,\mathrm{mm}$, $\theta_0=22.5°$) and transducer angle ($\theta_i=30°$) were adjusted to the experiments. The pipe was divided by eight quadratic cylindrical elements in the r direction as in Ref. 19. Calculations were done for 11 circumferential harmonics from $n=-5$ to 5 and for 1000 frequency steps from 0 to 2.0 MHz.

Figure 4 shows the comparisons between experimental and computated waveforms (displacements) when the inter-transducer distance L was 800 mm. Figure 4(a) presents the distribution of maximum amplitude of the waveforms detected at various circumferential positions (circumferential profile). Figure 4(b) shows waveforms obtained both experimentally and calculated at various circumferential positions. Since the waveforms at the circumferential position of 225°, 270°, and 315° were very similar to those at these symmetric positions 135°, 90°, and 45°, waveforms only at these five positions were shown in this figure. These SAFE calculations agree well with experiments not only in amplitude distribution and arrival time of waveforms, but also in phase, which proves that these calculations were carried out with sufficient accuracy.

Figure 5 shows circumferential profiles at a target position in the focusing experiments. Amplitudes and time delays were calculated for a target point of $L=400\,\mathrm{mm}$, 0° and of L=800 mm, 0° (Table I). Large signals were detected at the target point of 0° in the both cases. This shows that focusing phenomena occur at the target points given by the SAFE calculations and the focusing algorithm. Calculated circumferential profiles are also shown in Fig. 4, where the focusing phenomena were simulated using amplitudes and time delays, these results being visualized later. Main differences between experimental and calculated profiles were caused by the fact that little excitation signals from eight transducers were not all well-formed tone-burst identical signals. This hardware issue can be improved with little difficulty.

Furthermore, interesting knowledge is obtained by comparing the amplitudes and time delays given in Table I with the waveforms shown in Fig. 4. Amplitudes and time delays obtained from the waveforms in Fig. 5 are shown in Table II, where the waveform at 180° is assumed to be standard. By adding nT (n is integer and T is a period; 3.45 μs for 290 kHz) into these time delays, values in the lowest row in Table II were obtained. Amplitudes and time delays in Table I $L=800$ and Table II show very good agreement but signs of the time delays are opposite. This means that the focusing can be realized as following. First, waveforms are detected at the N receivers, where a single source transducer is installed at a target point. Then, N received signals are time reversed. In the next step, N receivers are used as N transmitters and time-reversed signals are excited at these transmitters. This results in the focusing at the target point where the single

TABLE II. Amplitude and time delay t_{delay} obtained by waveforms in Fig. 3 and time-delay shift $t_{\mathrm{delay}}\pm nT$.

	Top, 0°	45°	90°	135°	Bottom, 180°
Amplitude [-]	0.64	0.66	0.68	0.30	1.00
Time delay [μs] t_{delay}	−5.21	4.41	−1.30	1.98	0.00
$t_{\mathrm{delay}}\pm nT$ [μs]	−1.76	−2.49	−1.30	−1.52	0.00

transducer was installed first. This is nothing other than the application of ultrasonic time reversal.[22,23]

IV. VISUALIZATION OF FOCUSING TECHNIQUE

Since this focusing technique is based on the algorithm using complicated circumferential profiles, actual wave propagation status is hard to describe. For example, if focal points exist at unwanted points, they prevent us from determining the location of a crack. In order to apply the focusing technique to crack detection and defect evaluation, we need to comprehend wave motions and focusing status. In this section, the focusing technique was investigated by way of visualization results from SAFE calculations.

Figures 6(a) and (b) show the focusing status when amplitudes and time delays were tuned for focusing at $L=400$ mm, 0° and $L=800$ mm 0°, respectively. Visualization regions are from 20 mm, just after excitation region, to 1000 mm, where the grid width in the longitudinal direction is 10 mm. The shift of grid points describes amplitudes of the wave and the shade is normal stress in the longitudinal direction. Circle charts in these figures represent circumferential profiles at the maximum amplitude position. As amplitudes and time delays were calculated for the harmonic wave of 290 kHz, modes were selected by controlling incident angle, in this case L(0,2) and its neighbor F(n,2) modes, designed to focus to a focal point. However, it is possible to excite the other unwanted modes due to a frequency bandwidth of the incident wave with finite width transducers. Two wave packets can be seen in both Figs. 6(a) and (b), the faster one being the L(0,2) and F(n,2) modes and the second being the L(0,1) and F(n,1) modes. In both L=400 mm and L=800 mm, L(0,2)+F(n,2) modes are focused to the focal point as expected, while L(0,1)+F(n,1) modes are not.

Figures 6(a) and (b) reveal that the focusing phenomenon does not occur abruptly at the focal point, but goes through the process in which the first wave gradually increases as it approaches the focal point and decreases after the focal point. Considering the focusing algorithm, complex displacement amplitudes at the target position are described as $\mathbf{q}=[1\ 0\ 0,...,0]^{\mathrm{T}}$. This means that the focused waves propagate after focusing like those excited by a single transducer at the focal point. In the vicinity of the focal points, circumferential profiles shows that waves spread with distance, while in the far field, complicated circumferential profiles which cannot be predicted without the aid of calculations are shown in Fig. 6(a) $t=220\,\mu$s. As for the wave before the focusing points, the time-reversal idea is helpful in understanding that the wave propagation is roughly symmetric with respect to the focal point. As seen in Fig. 6(a) $t=100\,\mu$s and (b) $t=140\,\mu$s, circumferential profiles are similar to those for a single transmitter in the near field, that is, waves spread with distance from the target, but profiles became more complicated in far field as in Fig. 6(b) $t=180\,\mu$s. From a crack detection point of view, even if unexpected focusing phenomena occur in the far field from the designed focal point, they do not result in a wrong evaluation of defects since these focal points are not close to each other, anyway.

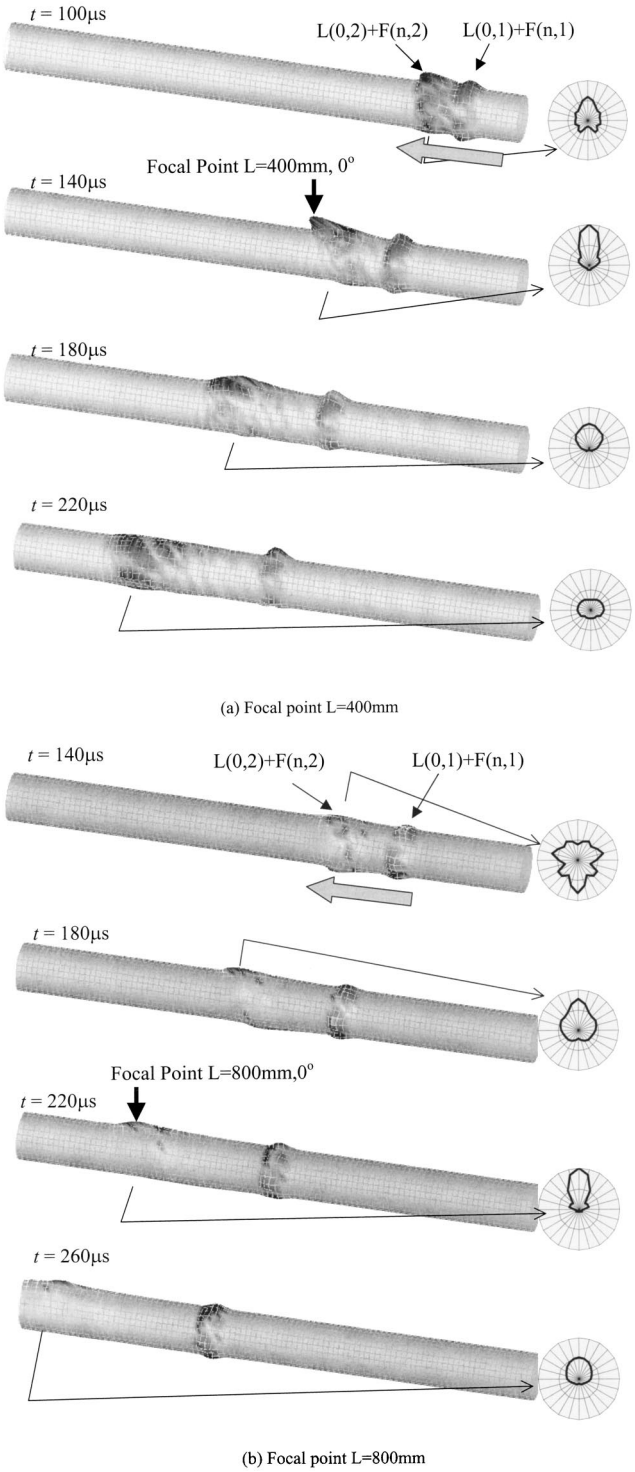


FIG. 6. Visualization results for focusing status. Wave propagation on the pipe surface and circumferential profiles at the location of L(0,2)+F(n,2) modes.

V. POTENTIAL OF THE FOCUSING TECHNIQUE

Focal spot size strongly affects the sensitivity of lateral resolution in crack detection. To make a focal spot small, the following ideas are considered: (1) to increase the number of channels; (2) to control incident waves by function generators; (3) to change the location of the transducers.

The focusing algorithm gives some explanations of the

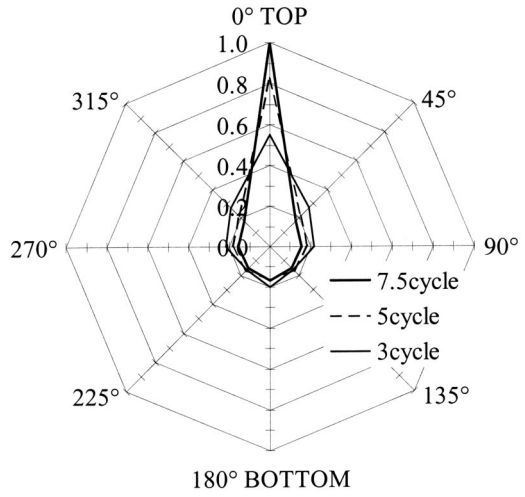


FIG. 7. Circumferential profiles for incident waves with different frequency bandwidth and cycles of tone-burst waves.

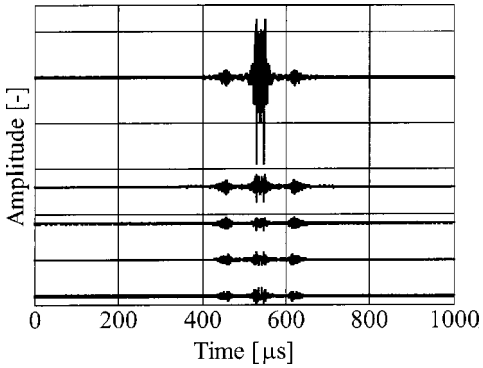


FIG. 8. Received signals for time-reversal emission tests (calculation) at the focal point (L=800 mm, 0°).

resolution with regard to the number of channels. As the number of channels N is large, the number of elements in vector $\mathbf{q}$ becomes large. This means that the circumferential width covered by one receiver becomes small, and this results in a small focusing spot. Detailed descriptions are written in Ref. 4.

Next, let us consider the control of incident waves. We discussed the use of tone-burst waves for focusing, in which amplitudes and time delays were obtained from a predicted circumferential profile by SAFE calculations only for a certain frequency ω (usually center frequency). However, if the circumferential profiles abruptly vary with a small shift of frequency $d\omega$, amplitudes and time delays obtained for a certain frequency ω are not designed to focus a wave of the frequency $\omega+d\omega$. Therefore, focus width could not be obtained as expected even in noise-less computer simulations. Figure 7 shows circumferential profiles at the focusing position with variable frequency bandwidth (number of cycles). As the number of cycles becomes small, that is, frequency bandwidth becomes broad, energy convergence at the focal point become small. This shows that amplitudes and time delays calculated for a center frequency do not work well for a converging ultrasonic wave at frequencies other than the center frequency.

For this reason, focusing is enhanced by applying amplitudes and time delays obtained independently for all frequency steps. Applying different amplitudes and time delays for different frequency steps means that arbitrary incident waves are excited instead of the tone-burst wave. From the time-reversal point of view,[21,22] this can be replaced as follows. N signals emitted from a single transducer and detected at N receivers are converted into N time-reversal waves. When the time-reversal waves are emitted back into the pipe from the N transducers, the ideal focusings that are not influenced by frequency bandwidth like that in Fig. 7 are realized. Figure 8 shows received waves (calculation) at L=800 mm when time reversal of calculated waveforms of the gray lines in Fig. 4 is used as excitation signals at eight transducers. Large waveforms were observed only at the desired focal point (L=800 mm, 0°), and small amplitudes at the other points.

Next, let us discuss the transducer location. The focusing algorithm can be applied even if transducers are not aligned in the circumferential direction. It is possible that changing the location of the transducers could improve focusing. However, since it is not easy to find the general algorithm on transducer location to optimize focusing, qualitative discussions will be described here. When the number of channels are N, the above-mentioned discussions are applied in this case, and the resolution does not change. But, the robustness of the solution $\mathbf{H}^{-1}$ is affected, and it results in a poor focusing characteristic. For example, when N transducers are at the same position, this is the same situation as a single transducer, and so focusing is not observed. Here, the matrix $\mathbf{H}$ is nonregular as $H_{11}=H_{21}=\cdots H_{N1}$, $H_{12}=H_{22}=\cdots H_{N2}$, $H_{13}=H_{23}=\cdots H_{N3},\ldots$, $H_{1N}=H_{2N}=\ldots H_{\mathrm{NN}}$. Generally speaking, the robustness of Eq. (14) depends on the independence of each equation in the simultaneous equations Eq. (14). In this sense, aligning transducers in the circumferential direction at a regular interval is the best arrangement to keep independence of these simultaneous equations and to give good focusing.

VI. CONCLUSIONS

Simulation of guided wave propagation was conducted using a semianalytical finite element method. This calculation was verified by way of a comparison of waveforms and an application to the focusing technique. Applying the semianalytical FEM calculations to focusing technique gave the following conclusions.

(1) Focusing does not occur abruptly at the focal point, but the amplitude of a wave packet increases as it approaches the focal point and decreases gradually after it. In the far field from the focal point, circumferential profiles become complicated. This is just like circumferential profiles from a single source at the focal point.
(2) Controlling incident waveforms, as well as increasing the number of channels, improves the resolution and sensitivity of the focusing technique. Using time-reversal signals realizes the ideal focusing in a pipe.

In this paper, circumferential resolutions in the focusing

technique were discussed, but it is necessary to investigate the resolution in the longitudinal direction, for which visualization will be helpful.

ACKNOWLEDGMENTS

We would like to thank Plant Integrity Ltd., Cambridge, England and Peter Mudge for their encouragement and partial support of this work on focusing in utilizing the nonaxisymmetric pipe modes.

[1] J. J. Ditri and J. L. Rose, ''Excitation of guided elastic wave modes in hollow cylinders by applied surface tractions,'' J. Appl. Phys. **72**(7), 2589–2597 (1992).
[2] H. J. Shin and J. L. Rose, ''Guided waves by axisymmetric and nonaxisymmetric surface loading on hollow cylinders,'' Ultrasonics **37**, 355–363 (1999).
[3] J. Li and J. L. Rose, ''Excitation and propagation of nonaxisymmetric guided waves in a hollow cylinder,'' J. Acoust. Soc. Am. **109**(2), 457–468 (2001).
[4] J. Li, ''Quantitative guided wave evaluation,'' Ph.D. thesis, The Pennsylvania State University (2001).
[5] J. L. Rose and X. Zhao, ''Flexural mode tuning for pipe elbow testing,'' Mater. Eval. **59**(5), 621–624 (2001).
[6] Z. Sun, J. L. Rose, M. Quarry, and D. Chinn, ''Flexural mode tuning in pipe inspection,'' in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 2002), Vol. 21, pp. 262–269.
[7] J. L. Rose, ''Guided Wave Ultrasonic Pipe Inspection—The Next Generation,'' Proceedings of the 8th ECNDT, Barcelona, Spain (2002).
[8] D. N. Alleyne, M. J. S. Lowe, and P. Cawley, ''The reflection of guided waves from circumferential notches in pipes,'' J. Appl. Mech. **65**, 635–641 (1998).
[9] R. Long, K. Vine, M. J. S. Lowe, and P. Cawley, ''Monitoring acoustic wave propagation in buried cast iron water pipes,'' in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 2001), Vol. 20, 1202–1209.
[10] P. Wilcox, M. Lowe, and P. Cawley, ''The effect of dispersion on long-range inspection using ultrasonic guided waves,'' NDT & E Int. **34**, 1–9 (2001).
[11] D. N. Alleyne, B. Pavlakovic, M. J. S. Lowe, and P. Cawley, ''Rapid, long-range inspection of chemical plant pipework using guided waves,'' in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 2001), Vol. 20, pp. 180–187.
[12] D. C. Gazis, ''Three-dimensional investigation of the propagation of waves in hollow circular cylinders. I. Analytical foundation. II. Numerical Results,'' J. Acoust. Soc. Am. **31**, 568–578 (1959).
[13] Y. K. Cheung, *Finite Strip Method in Structural Mechanics* (Pergamon, Oxford, 1976).
[14] O. C. Zienkiewicz and R. L. Taylor, *The Finite Element Method*, 5th ed. (Butterworth-Heinemann, Washington, D.C., 2000).
[15] S. K. Datta, A. H. Shah, R. L. Bratton, and T. Chakraborty, ''Wave propagation in laminated composite plates,'' J. Acoust. Soc. Am. **83**, 2020–2026 (1988).
[16] W. Karunasena, A. H. Shah, and S. K. Datta, ''Wave propagation in a multilayered laminated cross-ply composite plate,'' ASME Trans. J. Appl. Mech. **58**, 1028–1032 (1991).
[17] T. Kohl, S. K. Datta, and A. H. Shah, ''Axially symmetric pulse propagation in semi-infinite hollow cylinders,'' AIAA J. **30**(6), 1617–1624 (1992).
[18] N. Rattanawangcharoen, A. H. Shah, and S. K. Datta, ''Reflection of waves at the free edge of a laminated circular cylinder,'' J. Appl. Mech. **61**, 323–329 (1994).
[19] W. Zhuang, A. H. Shah, and S. B. Dong, ''Elastodynamic Green's function for laminated anisotropic circular cylinders,'' J. Appl. Mech. **66**, 665–674 (1999).
[20] H. Taweel, S. B. Dong, and M. Kazic, ''Wave reflection from the free end of a cylinder with an arbitrary cross section,'' Int. J. Solids Struct. **37**, 1701–1726 (2000).
[21] T. Hayashi and K. Kawashima, ''Multiple reflections of Lamb waves at a delamination,'' Ultrasonics **40**, 193–197 (2002).
[22] M. Fink, ''Time reversal of ultrasonic fields. I. Basic principles,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**(5), 555–566 (1992).
[23] R. K. Ing and M. Fink, ''Time-reversed Lamb waves,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**(4), 1032–1043 (1998).

Noncontact ultrasound stimulated optical vibrometry study of coupled vibration of arterial tubes in fluids

X. M. Zhang,[a] M. Fatemi, R. R. Kinnick, and J. F. Greenleaf
Ultrasound Research Laboratory, Mayo Clinic and Foundation, Rochester, Minnesota 55905

(Received 2 December 2001; revised 10 December 2002; accepted 22 December 2002)

Coupled vibration of arterial tubes is analyzed with the wave propagation approach and first-order shear deformation theory. Both the interior and exterior fluids are considered as compressible so that acoustic waves can be generated and propagated in the fluids. Results obtained using the theory have been evaluated against those available in the literature and the agreement has been found to be good. The theory can be used for future research on the vibration and acoustics of arterial walls. Vibration experiments were carried out on a silicone rubber tube in a water tank with a novel ultrasound stimulated optical vibrometry system. This system uses the radiation force of ultrasound to vibrate the tube at low frequency and records the resulting response by a laser vibrometer. Both the excitation and measurement are remote and noncontact. The silicone rubber tube was chosen because it has mechanical properties close to those of arteries. The fundamental frequency is well excited by the radiation force and measured with the laser. The measured fundamental frequency is in good agreement with the present theory.
[DOI: 10.1121/1.1548154]

PACS numbers: 43.20.Ks, 43.40.Ey, 43.80.Jz [DEC]

I. INTRODUCTION

The mechanical properties of the arterial wall are very important because they influence arterial physiology and the development and progression of arterial diseases via effects on blood flow and arterial mass transport. Furthermore, stresses and strains in the arterial wall are extremely important factors in understanding the pathophysiology and mechanics of the cardiovascular system. Stresses and strains cannot be analyzed without exact knowledge of the mechanical properties of the wall. Indeed, knowledge of the mechanical properties of the blood vessels has long been recognized as playing an essential role in cardiovascular behavior.[1–4] However, vibration analysis of arterial walls has not be extensively conducted.

This paper reports our recent study[5] on vibration of tubes. Here we study the coupled vibration of a tube immersed in and filled with fluid. We assume that: (a) the tube is of one layer made of homogeneous, isotropic, and elastic materials. The tube is considered as a thick shell, because the ratio of thickness to radius is relatively large for most arterial walls compared to thin shells widely used in engineering. (b) To simplify the problem, both the interior and exterior fluids are nonviscous; however, they are compressible, and therefore, acoustic waves can be generated and propagated in both fluids.

Vibration analyses of shells have been studied for over a century. Love[6] modified the Kirchhoff hypothesis for plates and established the preconditions of the so-called ''classic theory of thin shells,'' which is now commonly known as Love's first approximation of the first kind on which many thin shell theories were later developed, i.e., Donnell's theory.[7] Numerous methods have been developed and used to study the vibration behavior of thin shells based on these or other thin shell or classic theories. In classical or thin shell theories, which assume that straight lines normal to the middle surface before deformation remain straight, inextensible and normal to the middle surface after deformation, the transverse shear and transverse normal effects are ignored. This omission makes the thin shell theories highly inadequate for the analysis of even slightly thick shells. In recent years, the refinement of thin shell theories has resulted in a number of the so-called ''higher order shell theories.'' The first-order shear deformation theory has been used for vibration and dynamic stability analysis of thick shells.[8–10] An approximate analysis similar to the finite-strip method has been employed to study the vibration of thick cylindrical shells on the basis of three-dimensional theory of elasticity.[11]

If the tube vibrates in a vacuum or in a light fluid like air, the fluid loading effects on the tube can be ignored. Therefore, the vibration of the tube and acoustic waves in the fluid can be separately studied. Actually the coupled natural frequency of the tube in a light fluid is nearly the same as its uncoupled natural frequency. Therefore, no coupled analysis is needed, and calculations are significantly simplified. Numerous theories have been developed and used to study the vibration behavior of uncoupled shells, see Armenakas *et al.*[12] and Blevins.[13]

However, for an arterial tube that is in human tissues and carrying blood, the tissue, and blood introduce significant mass loading effects, and warrant a coupled fluid-structure analysis. It is then impossible to neglect the coupling between the tube with the exterior tissue and interior blood. In most cases it is not easy to exactly take into account the effect of the fluid-structure interaction by analytic methods, except in a few cases where an analytical solution can be made for simply shaped structures (e.g., spherical, or infinite cylindrical structures). Many studies have been made on fluid-loaded plate and shell problems in this field for the

[a]Author to whom correspondence should be addressed; electronic mail: zhang.xiaoming@mayo.edu

last few decades, see Morse *et al.*,[14] and Junger *et al.*[15] Zhang *et al.*[16] investigated the effects of baffles and end-caps on coupled structure-acoustic behaviors of finite cylindrical shells with numerical finite element method and boundary element method techniques. They concluded that the coupled frequencies were not influenced with the baffles and end-caps. It is possible to use a finite baffled cylindrical shell to model a finite cylindrical shell for coupled structure-acoustic analysis. Xu *et al.*[17] studied the effect of a wall joint on the vibrational power flow propagating in a fluid-filled shell, and the vibrational power flow from a line circumferential cosine harmonic force into an infinite elastic circular cylindrical shell filled with fluid.[18]

In this paper we present the theory for coupled vibration of arterial tubes using the wave propagation approach and first-order shear deformation theory. The wave propagation approach was proposed[19] for vibration analysis of thin cylindrical shells. The method has been extended for a thin cylindrical panel,[20] a thin laminated composite cylindrical shell,[21] and a thin fluid-filled pipe structure.[22] Similar work on the wave approach for vibration analysis has been presented by Wang *et al.*[23]

Experimental study on the coupled vibration of an arterial tube is quite different from those in engineering. Noncontact measurement is the basic requirement. Vibro-acoustography is a new noncontact imaging method based on the radiation force of ultrasound.[24–26] A comparative study of vibro-acoustography with other radiation force methods for tissue elasticity imaging has been presented.[27] The vibro-acoustography method is particularly useful in detecting hard inclusions in soft material, and has been used to image calcifications in human arteries.[27,28] In this study ultrasound stimulated displacement or velocity was measured with a laser vibrometer. We use our ultrasound stimulated laser vibrometry system for coupled vibration measurement of arterial tubes. As the first step in this kind of experiment, a thick silicone rubber tube was chosen for the experiment, because it has mechanical properties close to those of arterial tubes. The tube displacement was measured with a laser vibrometry system in a water tank. Also, both the interior and exterior fluids are water as a first approximation to tissue and blood. The measurement results are compared with the present theory.

II. THEORY

The cylindrical arterial tube under consideration has constant thickness h, radius R, and axial length L. The tube material is assumed to be homogeneous, isotropic, and linearly elastic with Young's modulus E, Poisson's ratio ν, and density ρ. The reference surface of the tube is taken to be at its middle surface where an orthogonal coordinate system (x, θ, z) is fixed. The x coordinate is taken in the axial direction of the tube, where the θ and z coordinates are, respectively, in the circumferential and radial directions of the tube as shown in Fig. 1. The tube is vibrating with both interior and exterior fluids. The interior fluid is designated with sound speed C_{fi} and mass density ρ_{fi}. The exterior fluid is expressed with sound speed C_{fe} and mass density ρ_{fe}.

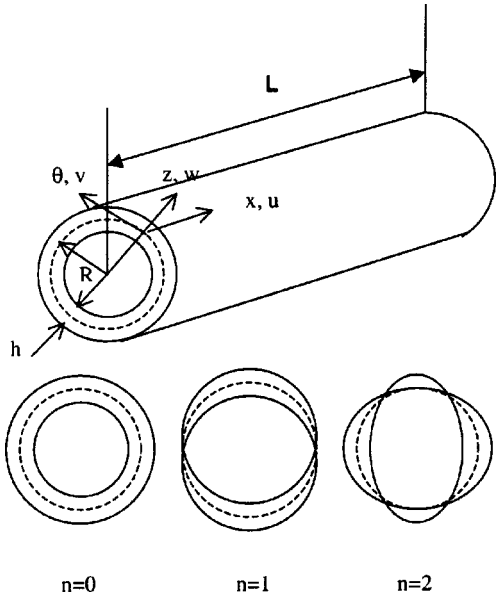


FIG. 1. Coordinate system and circumferential modal shapes.

A. Formulation

The deformations of the shell are defined by U, V, W, ϕ_x, and ϕ_θ, which are displacements of the point in the x, θ, z and the rotations of the transverse normal about the θ and x axis, respectively,

$$\partial U/\partial z=\phi_x, \quad \partial V/\partial z=\phi_\theta. \tag{1}$$

First-order shear deformation theory (FSDT)[8,9] is used to handle the effects of shear forces and rotary inertia on the thick shell. In the first-order shear deformation theory the Kirchhoff hypothesis is relaxed by not constraining the transverse normals to remain perpendicular to the midsurface after deformation. This amounts to including transverse shear strains in the theory. The inextensibility of transverse normals requires that W be independent of the thickness coordinate z. According to FSDT, the displacement fields are of the form

$$\begin{aligned} U(x,\theta,z,t)&=u(x,\theta,t)+z\phi_x(x,\theta,t),\\ V(x,\theta,z,t)&=v(x,\theta,t)+z\phi_\theta(x,\theta,t),\\ W(x,\theta,z,t)&=w(x,\theta,t), \end{aligned} \tag{2}$$

where u, v, w, ϕ_x, and ϕ_θ are the displacements of the point on the middle surface $z=0$ and the rotations of transverse normal about the θ and x axis, respectively.

The equations of free vibration with FSDT of a cylindrical tube using Donnell's theory can be written in terms of the forces and moment resultants,[9]

$$\begin{aligned} &\frac{\partial N_x}{\partial x}+\frac{1}{R}\frac{\partial N_{\theta x}}{\partial \theta}-I_0\frac{\partial^2 u}{\partial t^2}-I_1\frac{\partial^2 \phi_x}{\partial t^2}=0,\\ &\frac{\partial N_{x\theta}}{\partial x}+\frac{1}{R}\frac{\partial N_\theta}{\partial \theta}-I_0\frac{\partial^2 v}{\partial t^2}-I_1\frac{\partial^2 \phi_\theta}{\partial t^2}=0,\\ &\frac{\partial Q_x}{\partial x}+\frac{1}{R}\frac{\partial Q_\theta}{\partial \theta}-\frac{N_\theta}{R}-I_0\frac{\partial^2 w}{\partial t^2}=0, \end{aligned} \tag{3}$$

$$\frac{\partial M_x}{\partial x}+\frac{1}{R}\frac{\partial M_{\theta x}}{\partial \theta}-Q_x-I_1\frac{\partial^2 u}{\partial t^2}-I_2\frac{\partial^2 \phi_\theta}{\partial t^2}=0,$$

$$\frac{\partial M_{x\theta}}{\partial x}+\frac{1}{R}\frac{\partial M_\theta}{\partial \theta}+Q_\theta-I_1\frac{\partial^2 v}{\partial t^2}-I_2\frac{\partial^2 \phi_\theta}{\partial t^2}=0,$$

where N and M are the force and moment resultants defined as

$$\{N_x,N_\theta,N_{x\theta},N_{\theta x}\}=\int_{-h/2}^{h/2}\{\sigma_x(1+z/R),\sigma_\theta,\sigma_{x\theta}(1+z/R),\sigma_{x\theta}\}dz,$$

$$\{M_x,M_\theta,M_{x\theta},M_{\theta x}\}=\int_{-h/2}^{h/2}\{\sigma_x(1+z/R),\sigma_\theta,\sigma_{x\theta}(1+z/R),\sigma_{x\theta}\}z\,dz, \tag{4}$$

$$\{Q_x,Q_\theta\}=\int_{-h/2}^{h/2}\{\sigma_{xz}(1+z/R),\sigma_{\theta z}\}z\,dz$$

and the mass moments of inertia are

$$(I_0,I_1,I_2)=\int_{-h/2}^{h/2}\rho(1,z,z^2)dz.$$

For the cylindrical tube, the stresses defined in Eq. (4) are defined by the two-dimensional Hooke's law as

$$\begin{Bmatrix}\sigma_x\\ \sigma_\theta\\ \sigma_{\theta z}\\ \sigma_{xz}\\ \sigma_{x\theta}\end{Bmatrix}=\begin{bmatrix}Q_{11}&Q_{12}&&&\\ Q_{12}&Q_{22}&&&\\ &&Q_{44}&&\\ &&&Q_{55}&\\ &&&&Q_{66}\end{bmatrix}\begin{Bmatrix}e_x\\ e_\theta\\ e_{\theta z}\\ e_{xz}\\ e_{x\theta}\end{Bmatrix}. \tag{5}$$

The strain components in Eq. (5) are defined by[10]

$$\{e_x,e_\theta,e_{\theta z},e_{xz},e_{x\theta}\}=\left\{\frac{\partial U}{\partial x},\frac{1}{R}\left(\frac{\partial V}{\partial \theta}+W\right),\frac{\partial V}{\partial z}+\frac{1}{R}\frac{\partial W}{\partial \theta},\frac{\partial U}{\partial z}+\frac{\partial W}{\partial x},\frac{\partial V}{\partial x}+\frac{1}{R}\frac{\partial U}{\partial \theta}\right\}, \tag{6}$$

which can be written separately with FSDT as

$$\begin{Bmatrix}e_x\\ e_\theta\\ e_{\theta z}\\ e_{xz}\\ e_{x\theta}\end{Bmatrix}=\begin{Bmatrix}e_1\\ e_2\\ e_4\\ e_5\\ e_6\end{Bmatrix}+z\begin{Bmatrix}k_1\\ k_2\\ 0\\ 0\\ k_6\end{Bmatrix}, \tag{7}$$

where

$$\{e_1,e_2,e_4,e_5,e_6\}=\left\{\frac{\partial u}{\partial x},\frac{1}{R}\left(\frac{\partial v}{\partial \theta}+w\right),\phi_\theta+\frac{1}{R}\frac{\partial w}{\partial \theta},\phi_x+\frac{\partial w}{\partial x},\frac{\partial v}{\partial x}+\frac{1}{R}\frac{\partial u}{\partial \theta}\right\},$$

and

$$\{k_1,k_2,k_6\}=\left\{\frac{\partial \phi_x}{\partial x},\frac{1}{R}\frac{\partial \phi_\theta}{\partial \theta},\frac{\partial \phi_\theta}{\partial x}+\frac{1}{R}\frac{\partial \phi_x}{\partial \theta}\right\}.$$

The equations of free vibration of Donnell's classical shell theory are obtained from Eq. (3) by neglecting rotary inertia terms involving I_1 and I_2, and by setting $\phi_\theta=-(1/R)(\partial w/\partial \theta)$ and $\phi_x=-\partial w/\partial x$.

By substituting Eqs. (6) and (7) into Eq. (5), and then substituting the resulting equation into Eq. (4), the force and moment results can be derived as

$$\begin{Bmatrix}N_x\\ N_\theta\\ N_{x\theta}\\ N_{\theta x}\\ M_x\\ M_\theta\\ M_{x\theta}\\ M_{\theta x}\\ Q_x\\ Q_\theta\end{Bmatrix}=\begin{bmatrix}A_{11}&A_{12}&0&0&0&D_{11}/R&D_{12}/R&0\\ A_{12}&A_{22}&0&0&0&0&0&0\\ 0&0&0&0&A_{66}&0&0&D_{66}/R\\ 0&0&0&0&A_{66}&0&0&0\\ D_{11}/R&D_{12}/R&0&0&0&D_{11}&D_{12}&0\\ 0&0&0&0&0&D_{12}&D_{22}&0\\ 0&0&0&0&D_{66}/R&0&0&D_{66}\\ 0&0&0&0&0&0&0&D_{66}\\ 0&0&0&K_sA_{55}&0&0&0&0\\ 0&0&K_sA_{44}&0&0&0&0&0\end{bmatrix}\begin{Bmatrix}e_1\\ e_2\\ e_4\\ e_5\\ e_6\\ k_1\\ k_2\\ k_6\end{Bmatrix}, \tag{8}$$

where K_s is the shear correction factor, A_{ij}, D_{ij} are the extensional and bending stiffness defined, respectively, as

$$(A_{ij},D_{ij})=\int_{-h/2}^{h/2}Q_{ij}(1,z^2)dz. \tag{9}$$

For the isotropic tube, which is assumed to be in a state of plane stress, Q_{ij} are the reduced stiffness defined as

$$Q_{11}=\frac{E}{1-\nu^2},\quad Q_{12}=\frac{E\nu}{1-\nu^2},\quad Q_{22}=\frac{E}{1-\nu^2},$$
$$Q_{44}=\frac{E}{2(1+\nu)},\quad Q_{55}=\frac{E}{2(1+\nu)},\quad Q_{66}=\frac{E}{2(1+\nu)}, \tag{10}$$

where E and ν are the Young's modulus and Poisson's ratio for the tube's material.

By substituting Eq. (8), with substitutions from Eqs. (9) and (10), into Eq. (3), Eq. (3) can be written in a matrix form as

$$\begin{bmatrix} L_{11} & L_{12} & L_{13} & L_{14} & L_{15} \\ L_{21} & L_{22} & L_{23} & L_{24} & L_{25} \\ L_{31} & L_{32} & L_{33} & L_{34} & L_{35} \\ L_{41} & L_{42} & L_{43} & L_{44} & L_{45} \\ L_{51} & L_{52} & L_{53} & L_{54} & L_{55} \end{bmatrix} \begin{Bmatrix} u \\ v \\ w \\ \phi_x \\ \phi_\theta \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix}, \tag{11}$$

where L_{ij} $(i,j=1,2,3,4,5)$ are the differential operators with respect to x and θ.

The fluid interior or exterior to the tube satisfies the acoustic wave equation and the equation of motion of the fluid can be written in the cylindrical coordinate system (x, θ, r) as

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial p}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2 p}{\partial \theta^2}+\frac{\partial^2 p}{\partial x^2}=\frac{1}{C_f^2}\frac{\partial^2 p}{\partial t^2}, \tag{12}$$

where t is the time, p is the acoustic pressure, and C_f is the sound speed of the fluid. The x and θ coordinates are the same as those of the shell, where the r coordinate is taken from the x-axis of the shell.

B. Solution

In the research on vibration of arterial tubes, we need to calculate the resonant frequency of an arterial tube. Although blood flow and wave propagation in arteries have been extensively studied, the research on vibration of arterial tubes is relatively rare. This may be due to the fact that an artery is a continuous tree-like tube. It is difficult to find appropriate boundary conditions for an arterial branch. Therefore, it is very important for our solution to easily handle various boundary conditions of a thick tube.

For solution of resonant frequency of a thick tube we apply a method named wave propagation approach for frequency analysis of tubes because the method can easily treat various boundary conditions. In the wave propagation approach the displacements of the tube are expressed in the format of wave propagation, associated with an axial wave number k_x and circumferential modal parameter n, and defined by

$$\begin{aligned} u &= U_m \cos(n\theta)e^{(i\omega t-ik_x x)}, \\ v &= V_m \sin(n\theta)e^{(i\omega t-ik_x x)}, \\ w &= W_m \cos(n\theta)e^{(i\omega t-ik_x x)}, \\ \phi_x &= (\Phi_{xm}/R)\cos(n\theta)e^{(i\omega t-ik_x x)}, \\ \phi_\theta &= (\Phi_{\theta m}/R)\sin(n\theta)e^{(i\omega t-ik_x x)}, \end{aligned} \tag{13}$$

where U_m, V_m, and W_m are, respectively, the displacement amplitudes in the x, θ, z directions, Φ_{xm} and $\Phi_{\theta m}$ are, respectively, the rotation amplitudes of transverse normal about the θ and x axis, ω is the circular angular frequency.

The associated form of the acoustic pressure field in the interior fluid of tube, which satisfies the acoustic wave Eq. (12), can be expressed in the cylindrical coordinate system, associated with an axial wave number k_x, radial wave number k_{ri}, and circumferential modal parameter n, and given as[22]

$$p_i = P_{mi}\cos(n\theta)J_n(k_{ri}r)e^{(i\omega t-ik_x x)}, \tag{14}$$

where $J_n(\ )$ is the Bessel function of order n. k_x and k_{ri} are the axial and radial wave numbers, respectively. The radial wave number k_{ri} is related to the axial wave number k_x by the usual vector relation $(k_{ri}R)^2=\Omega^2(C_L/C_{fi})^2-(k_xR)^2$, where Ω is the nondimensional frequency, and C_L is the sound speed of the tube material.

The acoustic pressure field exterior to the shell, which also satisfies the acoustic wave Eq. (12), can be given as

$$p_e = P_{me}\cos(n\theta)H_n^{(2)}(k_{re}r)e^{(i\omega t-ik_x x)}, \tag{15}$$

where $H_n^{(2)}(\ )$ is the second kind Hankel function of order n. k_{re} is the radial wave number of exterior fluid, $(k_{re}R)^2=\Omega^2(C_L/C_{fe})^2-(k_xR)^2$.

To ensure that the fluid remains in contact with the tube wall, the fluid radial displacement and the tube radial displacement must be equal at the interface of the tube and the fluid. For the interior fluid this coupling condition requires

$$-\{1/(i\omega\rho_{fi})\}(\partial p_i/\partial r)\big|_{r=R_i}=(\partial w/\partial t)\big|_{r=R}, \tag{16}$$

and

$$P_{mi}=\{\omega^2\rho_{fi}/k_{ri}J_n'(k_{ri}R_i)\}W_m, \tag{17}$$

where the prime on the $J_n(\ )$ denotes differentiation with respect to the argument $k_{ri}R_i$.

Similarly, the coupling condition for the exterior fluid and the tube results in

$$P_{me}=\{\omega^2\rho_{fe}/k_{re}H_n'^{(2)}(k_{re}R_e)\}W_m, \tag{18}$$

where the prime denotes differentiation of the second Hankel function with respect to the argument $k_{re}R_e$, R_i, R_e are inner and outer radius, respectively.

Substituting Eq. (13) for Eq. (11), with consideration of acoustic pressures on both the interior and exterior walls of tube and the coupling Eq. (17) and Eq. (18), results in the equation of motion of coupled system in a matrix form as

$$\begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} \\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} \\ C_{31} & C_{32} & C_{33}+\mathrm{FL}_i-\mathrm{FL}_e & C_{34} & C_{35} \\ C_{41} & C_{42} & C_{43} & C_{44} & C_{45} \\ C_{51} & C_{52} & C_{53} & C_{54} & C_{55} \end{bmatrix} \begin{Bmatrix} U_m \\ V_m \\ W_m \\ \Phi_{xm} \\ \Phi_{\theta m} \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix}, \tag{19}$$

where C_{ij} $(i,j=1,2,3,4,5)$ are the parameters from the L_{ij} after they are operated with the x and θ. The parameters are listed in the Appendix of the paper.

FL_i is the interior fluid loading term due to the presence of the fluid acoustic field and is given by[22]

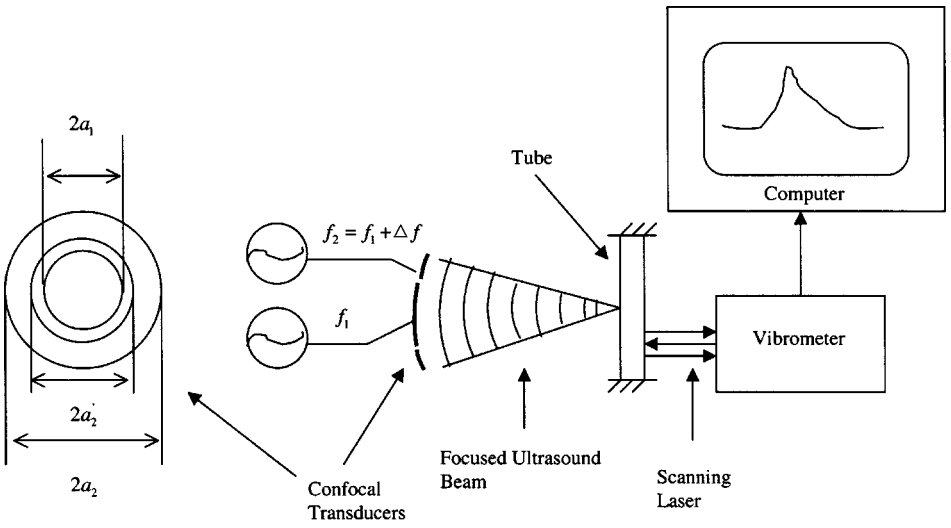


FIG. 2. Ultrasound stimulated optical vibrometry system. The system includes a focused confocal transducer. Two continuous-wave generators drive these elements at slightly different frequencies. The transducer is focused on the tube, with the beams interacting at the joint focal point to produce an oscillating radiation force on the tube at the difference frequency. This force causes the tube to vibrate. The response of the tube is measured with a VibraScan Laser Vibrometer.

$$\mathrm{FL}_i = \Omega^2(\rho_{fi}/\rho_s)(R/h)(k_{ri}R)^{-1}[J_n(k_{ri}R_i)/J_n'(k_{ri}R_i)]. \tag{20}$$

FL_e is the exterior fluid loading term given by

$$\mathrm{FL}_e = \Omega^2(\rho_{fe}/\rho_s)(R/h) \times (k_{re}R)^{-1}[H_n^{(2)}(k_{re}R_e)/H_n'^{(2)}(k_{re}R_e)]. \tag{21}$$

For nontrivial solutions, one sets the determinant of the characteristic matrix in Eq. (19) to zero. Expansion of the determinant of the above equation provides the system characteristic equation

$$F(k_x, \omega) = 0, \tag{22}$$

where $F(k_x, \omega)$ is a complex nonlinear and transcendental function, not a polynomial function as in the uncoupled situation. This characteristic function can be used to investigate the wave propagation in the tube as well as the natural frequency of a finite tube.

Equation (22) is used here to obtain the coupled natural frequencies of the tube with interior and exterior fluids. In the wave propagation approach the wave number k_x is given according to appropriate boundary conditions of the tube, Eq. (22) can then be written as

$$P_1(\omega) - P_2(\omega)(\mathrm{FL}_i - \mathrm{FL}_e) = 0, \tag{23}$$

where both $P_1(\omega)$ and $P_2(\omega)$ are polynomial functions. For the sake of brevity, the coefficients are not given here.

In the case of uncoupled analysis, $\mathrm{FL}_i = \mathrm{FL}_e = 0$, Eq. (23) can be written as

$$P_1(\omega) = \omega^{10} + \beta_1\omega^8 + \beta_2\omega^6 + \beta_3\omega^4 + \beta_4\omega^2 + \beta_5 = 0, \tag{24}$$

where β_i $(i = 1,2,3,4,5)$ are the coefficients of Eq. (24). Solving Eq. (24), one obtains five pairs of roots. The lowest of the positive roots represents the flexural vibration; the other four roots are two in-plane vibrations and two rotation vibrations.

For coupled analysis, Eq. (23) is a nonlinear and transcendental equation for either interior fluid or exterior fluid, or both interior and exterior fluids. The roots ω can be obtained by using Newton's procedure with starting values of ω_0. These starting values of ω_0 can be taken from the corresponding uncoupled frequencies.

The right axial wave number k_x must be chosen to satisfy the required boundary conditions at the two ends of the cylindrical tube, for the frequency to be obtained from Eq. (23). In the wave propagation approach the wave traveling in the axial direction of the tube is approximately obtained by studying the wave traveling in a similar beam. For the simply supported–simply supported (SS-SS) boundary conditions, $k_x = m\pi/L$ and n can be used for Eq. (23) to solve the natural frequency of the tube for the modal parameters (m, n). For a tube with the clamped–clamped (CC) boundary conditions, $k_x = (2m+1)\pi/(2L)$ and n should be used to find the frequency of the CC tube for the modal parameters (m, n). For other kinds of boundary conditions, the wave number equations are not so simple as that for the SS-SS boundary conditions. The equations are transcendental ones. Therefore there are not simple formulas for the wave numbers. However approximate wave numbers were proposed for other boundary conditions in Ref. 22.

III. MEASUREMENTS

The ultrasound stimulated optical vibrometry system uses radiation force to vibrate the object. Object vibrations are measured with a laser vibrometer. To generate a localized oscillatory stress field, two intersecting cw focused ultrasound beams of different frequencies are used. It is only in the intersection region that the ultrasound field energy density is sinusoidally modulated, and hence, the field generates an oscillatory radiation force, which can interact with the object.

The experimental setup is shown in Fig. 2. The experiments were conducted in a water tank. The silicone rubber tube was placed at the focal plane of the ultrasound beams in a water tank. Water temperature was 21.2 °C (not controlled). The focused transducer is with $a_2 = 22.5$ mm, $a_1 = 14.8$ mm, and $a_2' = 16.8$ mm. The focal distance is $z_0 = 70$ mm, and the center frequency, f_1, was 3 MHz. Transducer elements were driven by two stable rf synthesizers (HP

TABLE I. Comparison of the frequency parameter $\Omega = (\omega h/\pi)\sqrt{2(1+\nu)\rho/E}$ for a simply supported–simply supported cylindrical shell ($m=1$, $n=1$, $h/R=0.05$, $\nu=0.3$).

h/L	Armenakas *et al.*[a]	Present
0.1	0.029 17	0.029 213
0.2	0.062 93	0.062 874
0.3	0.123 66	0.123 154
0.4	0.200 49	0.199 09
0.5	0.287 16	0.284 391
0.6	0.379 86	0.375 285
0.7	0.476 21	0.469 462
0.8	0.574 75	0.565 502
0.9	0.674 51	0.662 53
1.0	0.774 91	0.760 007

[a]Reference 12.

33120A) at frequencies of $f_1=3$ MHz and $f_2=f_1+\Delta f$, where Δf was swept slowly from 50 to 1000 Hz.

The vibration responses of the tube were measured with a Polytec VibraScan Laser Vibrometer system. The system consists of an OFV056 Vibrometer Scanning Head and an OFV 3001S Vibrometer Controller. With this system the objects can be scanned automatically at predescribed points. In our experiment, a scan along the axial direction of the tube was automatically done with the Polytec VibraScan system. Fifteen measurement points were used for the axial scanning along the tube length.

IV. RESULTS AND DISCUSSION

A. Validation

The accuracy of the wave propagation approach for uncoupled frequencies of shell has been validated previously in Ref. 5. In this paper more comparison results are given for further validations. For shells without fluid, the results can be compared with available published results in the literature. Table I shows the comparison for a thin simply supported–simply supported (SS–SS) shell against results presented by Armenakas *et al.*,[12] who used three-dimensional elasticity theory for the shell. The frequency parameter is $\Omega = (\omega h/\pi)\sqrt{2(1+\nu)\rho/E}$. The shell parameters are $h/R=0.05$, $m=1$, and $n=1$. Ten parameters of thickness-to-length h/L are considered. h/L is from 0.1 to 1.0. A similar comparison is given in Table II for a thicker shell with $h/R=0.35$.

TABLE II. Comparison of the frequency parameter $\Omega = (\omega h/\pi)\sqrt{2(1+\nu)\rho/E}$ for a simply supported–simply supported cylindrical shell ($m=1$, $n=1$, $h/R=0.35$, $\nu=0.3$).

h/L	Armenakas *et al.*[a]	Present
0.1	0.059 38	0.061 981
0.2	0.131 13	0.133 559
0.3	0.191 01	0.193 379
0.4	0.255 54	0.256 953
0.5	0.330 86	0.330 581
0.6	0.415 07	0.412 58
0.7	0.505 25	0.500 198
0.8	0.599 23	0.591 35
0.9	0.695 56	0.684 658
1.0	0.793 31	0.779 241

[a]Reference 12.

TABLE III. Comparison of frequency in hertz for an uncoupled clamped–clamped cylindrical shell between Love's theory and first shear deformation theory ($L=20$ m, $R=1$ m, $h=0.01$ m).

Order	FEM[a]	Love's[a]	Present	Modal shape (m, n)
1	12.25	12.17	13.51	(1,2)
2	19.64	19.61	21.92	(1,3)
3	23.18	23.28	25.29	(2,3)
4	27.69	28.06	28.71	(2,2)
5	31.6	31.98	33.50	(3,3)
6	36.7	36.47	38.88	(1,4)
7	37.55	37.37	39.75	(2,4)
8	39.87	39.78	42.05	(3,4)

[a]Adapted from Zhang *et al.*, Ref. 18.

Very few results can be found for coupled natural frequencies of cylindrical shell in the published literature. The accuracy of the wave propagation approach for a thin cylindrical shell filled with still fluid was validated in Ref. 22 with the Love's thin theory against the results of numerical FEM/BEM. The results are used for validation of the present results with the first-order shear deformation theory. Table III shows the comparison of natural frequencies in hertz with Zhang *et al.*[18] for a corresponding uncoupled clamped–clamped (CC) cylindrical shell. The coupled frequencies are compared in Table IV. It can be seen from the comparison that good agreement is found between Love's theory and first-order shear deformation theory.

B. Experimental results

Experimental studies were carried out on a thick silicone rubber tube. The tube ends were fixed with interior steel rods and clamped on the exterior surface at both ends. However the simulation and experimental modal shape show that the boundary conditions are simply supported. The tube was filled with water and immersed in a water tank. The detailed experimental setup for the underwater tube is shown in Fig. 3. The tube is painted on the surface so the laser reflects well. The confocal transducer is behind the tube in the picture. The length, outside diameter, inside diameter, and wall thickness of the tube are, respectively, 30 mm, 6.35 mm, 3.18 $\pm$0.30 mm, and 1.59$\pm$0.15 mm. The central radius R is 2.38

TABLE IV. Comparison of frequency in hertz for a coupled clamped–clamped cylindrical shell between Love's theory and first shear deformation theory ($L=20$ m, $R=1$ m, $h=0.01$ m).

Order	FEM/BEM[a]	Love's[a]	Present	Modal shape (m, n)
1	4.89	4.93	5.49	(1,2)
2	9.00	8.94	9.99	(1,3)
3	10.64	10.64	11.55	(2,3)
4	11.12	11.48	11.74	(2,2)
5	14.55	14.66	15.35	(3,3)
6	18.55	18.26	19.45	(1,4)
7	19.00	18.73	19.91	(2,4)
8	20.21	19.96	21.10	(3,4)

[a]Adapted from Zhang *et al.*, Ref. 22.

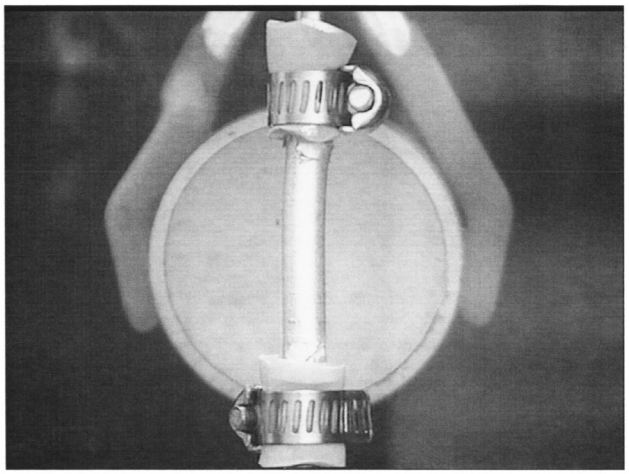
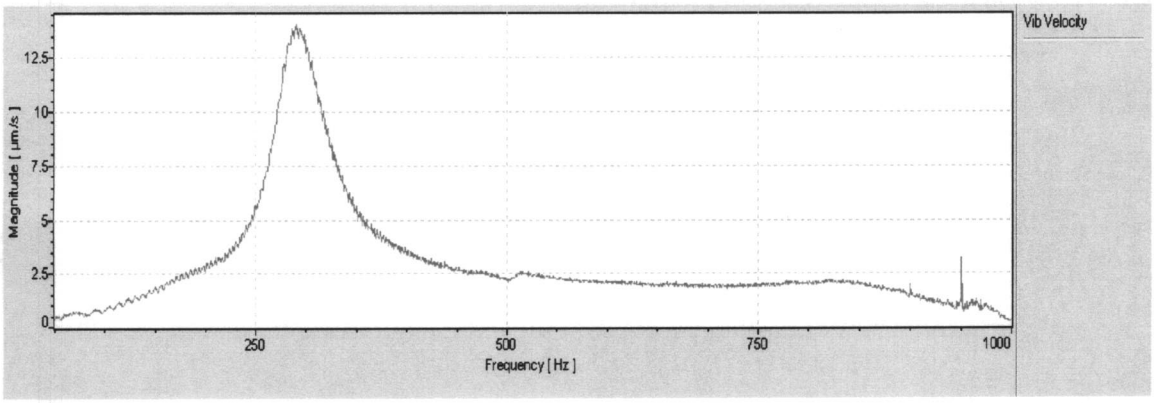


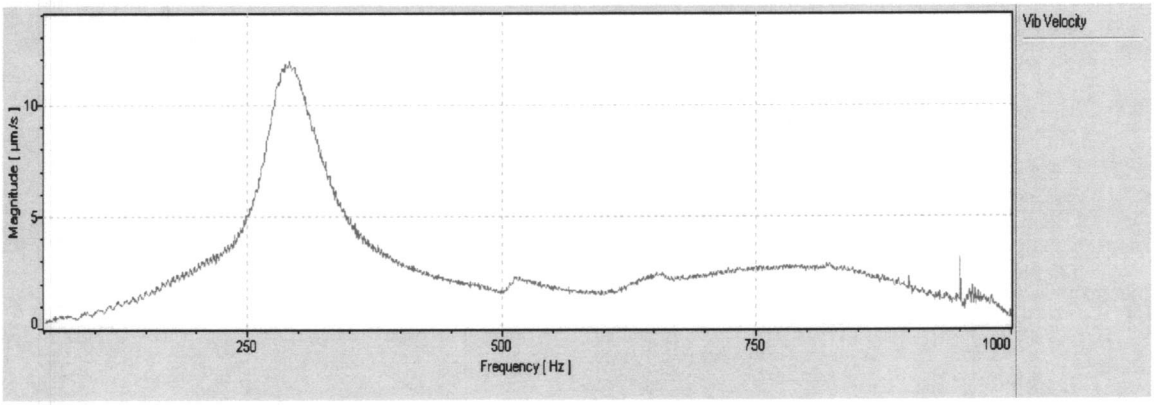
FIG. 3. Detailed experimental setup for an underwater tube.

mm. The ratio of thickness to radius is about 0.66, so the tube should be considered as a thick shell. The thin shell theories cannot be used for theoretical prediction in this case. The elastic parameters of the tube material are: mass density $\rho = 1180\ \text{kg/m}^3$, Poisson's ratio $\nu = 0.44$, and Young's modulus $E = 2.1 \times 10^6\ \text{N/m}^2$.

The tube was excited with the radiation force of ultrasound and the displacement velocity was measured with the Laser Vibrometer system. The frequency range of interest was from 50 to 1000 Hz. The temporal fast Fourier transform sample number was 1600, resulting in a frequency resolution of 0.625 Hz. The velocity frequency responses are shown at the point $(x_r, \theta) = (x_r, \pi)$, which is on the opposite side of the excitation point $(x_i, \theta) = (x_i, 0)$ in terms of circumferential direction. x_r and x_i are taken from the bottom of the tube. $x_i = 15$ mm means the focused point is at the center of the tube. Figure 4 shows the frequency responses at points $(x_r, \pi) = (15\ \text{mm}, \pi)$ and $(17\ \text{mm}, \pi)$ in terms of the symmetric excitation of $(x_i, 0) = (15\ \text{mm}, 0)$. Figure 5 shows the frequency responses at the same points $(x_r, \pi) = (15\ \text{mm}, \pi)$ and $(17\ \text{mm}, \pi)$ but in terms of the asymmetric excitation of $(x_i, 0) = (22.5\ \text{mm}, 0)$. In the frequency range of interest the fundamental frequency can be identified clearly either under symmetric or asymmetric excitations. However, higher order resonant frequencies cannot be identified clearly in the frequency region. This may due to the small radiation force and high damping of the tube. In most cases the fundamental frequency is the most important frequency and contains information about the elastic properties

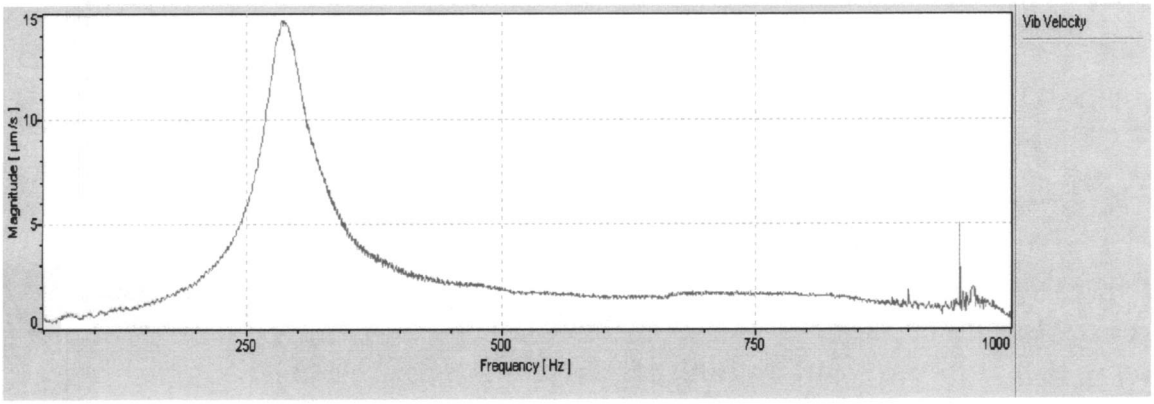


(a)

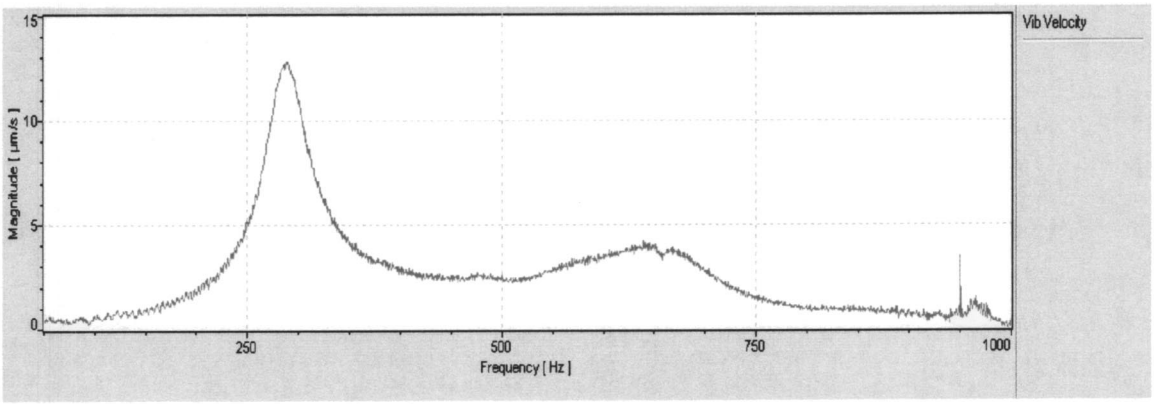


(b)

FIG. 4. Velocity frequency responses at points (a) $(x_r, \pi) = (15\ \text{mm}, \pi)$, (b) $(17\ \text{mm}, \pi)$ under the symmetric excitation at $(x_i, 0) = (15\ \text{mm}, 0)$, measured with laser vibrometry.



(a)



(b)

FIG. 5. Velocity frequency responses at points (a) $(x_r, \pi) = (15\text{ mm}, \pi)$, (b) (17 mm, π) under the asymmetric excitation at $(x_i, 0) = (22.5\text{ mm}, 0)$, measured with laser vibrometry.

of the material, therefore, it is very important to experimentally determine the fundamental frequency. The fundamental frequency can be found in Fig. 4 or 5 and is 291.3 Hz. In order to compare with the theoretical analysis, the fundamental frequency is calculated with the theory developed in this paper. The calculated fundamental frequency is 295.0 Hz. It can be seen that the experimental results agree fairly well with the theoretical prediction.

V. CONCLUSION

In this paper we present a theory for coupled vibration analysis of arterial tubes using the wave propagation approach and first-order shear deformation theory. An artery-like tube is modeled with thick shell theory. Both the interior and exterior fluids are considered as nonviscous but compressible. Our coupled vibration analysis has been validated with published results and experiments and good agreement has been found. Further studies will be made when more mechanical properties of the arterial tubes, tissue, and blood are taken into consideration.

Vibration experiments were carried out on a silicone rubber tube with ultrasonic stimulated optical vibrometry, which uses the radiation force of ultrasound to vibrate the tube at low frequency and the resulting displacement or velocity response were recorded with a laser vibrometer. The measured fundamental resonance frequency is found to be in good agreement with that of the theory. Both excitation and measurement were remote and noncontact, important attributes for future vibration measurements of arteries.

ACKNOWLEDGMENTS

This study is supported by Grant No. HL 61451 from the National Institutes of Health. The authors thank Thomas M. Kinter for software support and Elaine C. Quarve for secretarial assistance.

APPENDIX

The coefficients C_{ij} $(i,j = 1,2,3,4,5)$ in Eq. (19) are defined as follows:

$$C_{11} = -A_{11}k_x^2 - A_{66}n^2/R^2 + I_0\omega^2,$$

$$C_{12} = -iA_{12}k_x n/R - iA_{66}k_x n/R, \quad C_{13} = -iA_{12}k_x/R,$$

$$C_{14} = -D_{66}n^2/R^4 - D_{11}k_x^2/R^2 + I_1\omega^2/R,$$

$$C_{15}=-iD_{12}k_x n/R^3-iD_{66}k_x n/R^3,$$

$$C_{21}=iA_{12}k_x n/R+iA_{66}k_x n/R,$$

$$C_{22}=-A_{66}k_x^2-A_{22}n^2/R^2+I_0\omega^2, \quad C_{23}=-A_{22}n/R^2,$$

$$C_{24}=iD_{66}k_x n/R^3, \quad C_{25}=-D_{66}k_x^2/R^2+I_1\omega^2/R,$$

$$C_{31}=iA_{12}k_x/R, \quad C_{32}=-iA_{22}n/R^2,$$

$$C_{33}=-A_{55}K_5k_x^2-A_{22}/R^2-A_{44}K_4n^2/R^2+I_0\omega^2,$$

$$C_{34}=-iA_{55}K_5k_x/R, \quad C_{35}=A_{44}K_4n/R^2,$$

$$C_{41}=-D_{11}k_x^2/R+I_1\omega^2, \quad C_{42}=-iD_{12}k_x n/R^2,$$

$$C_{43}=iA_{55}K_5k_x-iD_{12}k_x/R^2,$$

$$C_{44}=-D_{66}n^2/R^3-A_{55}K_5/R-D_{11}k_x^2/R+I_2\omega^2/R,$$

$$C_{45}=-iD_{12}k_x n/R^2-iD_{66}k_x n/R^2,$$

$$C_{51}=iD_{66}k_x n/R^2, \quad C_{52}=-D_{66}k_x^2/R+I_1\omega^2,$$

$$C_{53}=A_{44}K_4n/R, \quad C_{54}=iD_{12}k_x n/R^2+iD_{66}k_x n/R^2,$$

$$C_{55}=-D_{22}n^2/R^3-A_{44}K_4/R-D_{66}k_x^2/R+I_2\omega^2/R,$$

where ω is the frequency, R the radius, n the circumferential mode, $K_4=K_5=K_s=5/6$ the shear correction factor, and k_x the axial wave number.

[1] W. W. Nichols and M. F. O'Rourke, *McDonald's Blood Flow in Arteries*, 4th ed. (Arnold, London, 1998).
[2] J. D. Bronzino, *The Biomedical Engineering Handbook* (CRC Press, Boca Raton, FL, 1995).
[3] G. W. Hastings, *Cardiovascular Biomaterials* (Springer, London, 1992).
[4] C. Leondes, *Cardiovascular Techniques* (CRC Press, Boca Raton, FL, 2001).
[5] X. M. Zhang, J. F. Greenleaf, M. Fatemi, and R. R. Kinnick, "A preliminary vibration model for frequency analysis of arterial walls," J. Sound Vib. (submitted).
[6] A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity*, 4th ed. (Cambridge University Press, Cambridge, 1952).
[7] L. H. Donnell, "Stability of thin walled tubes under torsion," NACA Report No. 479, 1993.
[8] A. Nosier and J. N. Reddy, "Vibration and stability analyses of cross-ply laminated circular cylindrical shells," J. Sound Vib. **157**, 139 (1992).
[9] T. Y. Ng, K. Y. Lam, and J. N. Reddy, "Dynamic stability of cylindrical panels with transverse shear effects," Int. J. Solids Struct. **36**, 3483–3496 (1999).
[10] K. Y. Lam and W. Qian, "Vibrations of think rotating laminated composite cylindrical shells," J. Sound Vib. **225**, 483–501 (1999).
[11] C. T. Loy and K. Y. Lam, "Vibrations of thin cylindrical shells on the basis of three-dimensional theory of elasticity," J. Sound Vib. **226**, 719–737 (1999).
[12] A. E. Armenakas, D. Gazis, and G. Herrmann, *Free Vibration of Circular Cylindrical Shells* (Pergamon, Oxford, 1969).
[13] R. D. Blevins, *Formulas for Natural Frequency and Mode Shape* (Van Nostrand Reinhold, New York, 1979).
[14] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw–Hill, New York, 1968).
[15] M. C. Junger and D. Feit, *Sound Structures, and Their Interaction* (Acoustical Society of America, Woodbury, New York, 1993).
[16] X. M. Zhang, G. R. Liu, and K. Y. Lam, "The effects of baffles and end-caps on coupled vibration and sound radiation of finite cylindrical shells," Int. J. Eng. Simulation **2**, 19–25 (2001).
[17] M. B. Xu and X. M. Zhang, "Vibrational power flow in a fluid-filled cylindrical shell," J. Sound Vib. **218**, 587–598 (1998).
[18] M. B. Xu, X. M. Zhang, and W. H. Zhang, "The effect of wall joint on the vibrational power flow propagation in a fluid-filled shell," J. Sound Vib. **224**, 395–410 (1999).
[19] X. M. Zhang, G. R. Liu, and K. Y. Lam, "Vibration analysis of thin cylindrical shells using wave propagation approach," J. Sound Vib. **239**, 397–403 (2001).
[20] X. M. Zhang, G. R. Liu, and K. Y. Lam, "Frequency analysis of cylindrical panels using a wave propagation approach," Appl. Acoust. **62**, 527–543 (2001).
[21] X. M. Zhang, "Vibration analysis of cross-ply laminated composite cylindrical shells using Wave propagation approach," Appl. Acoust. **62**, 1221–1228 (2001).
[22] X. M. Zhang, G. R. Liu, and K. Y. Lam, "Coupled vibration analysis of fluid-filled cylindrical shells using the wave propagation approach," Appl. Acoust. **62**, 229–243 (2001).
[23] C. Wang and J. C. S. Lai, "Prediction of natural frequencies of finite length circular cylindrical shells," Appl. Acoust. **59**, 385–400 (2000).
[24] M. Fatemi and J. F. Greenleaf, "Ultrasound stimulated vibro-acoustic spectroscopy," Science **280**, 82–85 (1998).
[25] M. Fatemi and J. F. Greenleaf, "Vibro-acoustography: An imaging modality based on ultrasound stimulated acoustic emission," Proc. Natl. Acad. Sci. U.S.A. **96**, 6603–6608 (1999).
[26] M. Fatemi and J. F. Greenleaf, "Application of radiation force in noncontact measurement of the elastic parameters," Ultrason. Imaging **21**, 147–154 (1999).
[27] M. Fatemi and J. F. Greenleaf, "Probing the dynamics of tissue at low frequencies with the radiation force of ultrasound," Phys. Med. Biol. **45**, 1449–1464 (2000).
[28] M. Fatemi and J. F. Greenleaf, "Imaging and evaluating the elastic properties of biological tissues," BMUS Bull. **8**, 16–18 (2000).

The scattering of guided waves in partly embedded cylindrical structures

T. Vogt, M. Lowe, and P. Cawley[a)]
Department of Mechanical Engineering, Imperial College, London SW7 2BX, United Kingdom

(Received 15 May 2002; revised 8 November 2002; accepted 12 November 2002)

The scattering of ultrasonic guided waves at a point where a free cylindrical waveguide enters an embedding material is investigated. A modal solution that is valid when the guided waves are incident from the free section of the waveguide is developed. It is shown that in this case it is valid to consider only the modal fields over the cross section of the waveguide, neglecting the fields in the embedding material. As an application, the scattering of the lowest-order longitudinal mode in a cylindrical waveguide, L(0,1), is examined in detail. As well as considering epoxy resin as an embedding material, the case where the embedding material is replaced by a perfectly rigid boundary is discussed. The latter gives some insight into the role of nonpropagating and inhomogeneous waveguide modes in the scattering process. The results from the modal solution are validated using Finite Element modeling, very good agreement being obtained. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1553463]

PACS numbers: 43.20.Mv, 43.35.Zc [ANN]

I. INTRODUCTION

The motivation for the work presented in this paper originates from an investigation into the possibility of using guided ultrasonic waves in a rod partially embedded in an epoxy resin for monitoring the cure process of the resin.[1] Two effects influence guided wave propagation when a waveguide is partly embedded in another material. First, the guided wave will be attenuated due to the leakage of bulk waves, depending on the material properties of the waveguide and the embedding material. Second, because of the change in surface impedance at the point where the waveguide enters the embedding material, the guided wave will be reflected and also scattered through mode conversion. Both these effects can be used to monitor the curing process, and the experimental techniques were first discussed by Papadakis,[2] but the results presented were of a qualitative nature. In order to make quantitative predictions, the guided wave attenuation can, in general, be calculated if, the acoustic properties of all the materials in the guided wave system are known;[3–5] indeed, the attenuation is an integral part of the dispersion curve calculation. To predict the scattering of a guided wave that is incident from the free section of the waveguide at a change in surface impedance is more difficult, and this problem is addressed in this paper.

Apart from measuring material properties by embedding a waveguide in the material under investigation (see, for example, Refs. 6–8), there are many other circumstances when the application of a guided wave system to solve a practical problem involves a waveguide that is partly free and partly embedded. Examples include the testing of partly buried pipes for corrosion[9] or the testing of rock bolts for defects in mining roofs from their protruding end.[10] In all these cases, depending on the waveguide mode, its frequency and the geometry of the waveguide, reflection will occur at the entry point. In many cases, a reflection from the entry point is undesirable, because it either interferes with other signals of interest, or can even make it impossible to test for defects in the embedded section if the reflection coefficient is so high that only little energy is transmitted. However, there are circumstances in which the presence of an entry reflection is advantageous. Examples are, as mentioned above, the cure monitoring of solidifying materials, or liquid level measurements.[11]

Most of these applications usually involve the use of a cylindrical waveguide. Some theoretical work has been done on scattering from discontinuities in single and layered plates,[12–15] but less attention has been given to scattering problems in cylindrical waveguides. Zemanek[16] treated the case of guided waves being scattered from the end of a cylinder; the scattering of guided waves from circumferential cracks in hollow cylinders was presented by Ditri[17] and Bai *et al.*[18] Engan[19] examined the scattering of torsional waves from a step change in the waveguide diameter. The theoretical work in this paper is based on some of the techniques presented in these references.

Finite Element (FE) modeling presents another possibility of solving scattering problems and has been employed by several researchers.[20–22] This is most useful when analytical solutions cannot be found. However, the computational analysis can be tedious and does not easily yield insight into the physics of the scattering process. Moreover, numerical problems occur when dealing with materials that have small shear velocities, for example, epoxy resins at the beginning of cure. It is therefore desirable to find an alternative solution for the cases where FE modeling is slow or not feasible. In the following, FE modeling will be used for the validation of the theoretical results where the material parameters allow for this.

First, a theoretical analysis will show the equivalence between the S-parameter formalism introduced by Auld[12] and a solution that is based on matching the fields at the interface between the free and the embedded sections of a cylindrical waveguide when a guided wave is incident from

a)Electronic mail: p.cawley@ic.ac.uk

 0001-4966/2003/113(3)/1258/15/$19.00

the free section. A complete modal solution is then given, and the results will be compared to FE modeling. It will be shown that in most cases of interest, only a small number of waveguide modes has to be considered in the modal solution. The following theoretical analysis is completely general, but for the purpose of demonstrating its validity and in view of the practical application presented in Ref. 1, the examples given will be restricted to the case where the lowest-order longitudinal mode is incident.

II. THEORETICAL BASIS

A. Background

For the following derivations it is necessary to introduce the notation used to describe guided modes in cylindrical structures. Time-harmonic propagation in the direction of the axis of the waveguide is considered, which is chosen to be the z direction in the cylindrical coordinates (r,Θ,z). A guided wave solution is represented by sets of wave number and frequency combinations for which the particle velocity and stresses satisfy the chosen boundary conditions at the surfaces of the cylindrical structure.[16,23] The particle velocity, $\mathbf{v}$, and stress tensor, $\mathbf{T}$, are given by the expressions

$$\mathbf{v}_n^\nu(r,\theta,z,t)=\mathbf{v}_n^\nu(r)e^{i\nu\theta}e^{-i(k_nz-\omega t)}, \tag{1}$$

$$\mathbf{T}_n^\nu(r,\theta,z,t)=\mathbf{T}_n^\nu(r)e^{i\nu\theta}e^{-i(k_nz-\omega t)}, \tag{2}$$

where ω is the circular frequency, k_n is the complex wave number in the propagation direction, and ν is the angular wave number. First, because of the continuity of the field variables in the angular direction, ν must be a whole number; it is commonly referred to as the circumferential order. For example, longitudinal and torsional modes are of circumferential order zero, which means that the modal fields are axisymmetric. For flexural modes, $\nu\geqslant 1$. Second, at each frequency, there is an infinite number of modes of a given circumferential order, which are solutions to the dispersion equation. The counter n is used to label a single waveguide mode of a given circumferential order. Therefore, the notation (ν,n) denotes the nth mode of circumferential order ν. These two indices completely identify a waveguide mode.

This notation is similar to that originally introduced by Silk and Bainton.[24] In fact, for flexural modes, the notation (ν,n) is equivalent to the bracketed term in the notation $\mathrm{F}(\nu,n)$ used by Silk and Bainton. However, for circumferential order zero the notation (ν,n) does not distinguish between longitudinal, L, and torsional, T, modes. In order to relate the counting number n used in this paper to the notation of Silk and Bainton in the case where $\nu=0$, the following convention can be made:

$n=1\rightarrow \mathrm{L}(0,1)$,

$n=2\rightarrow \mathrm{T}(0,1)$,

$n=3\rightarrow \mathrm{L}(0,2)$,

$n=4\rightarrow \mathrm{T}(0,2)$,

$n=5\rightarrow \mathrm{L}(0,3)$,

$\cdots$

Both of the above notations will be used where appropriate. The variation of the modal fields in the plane perpendicular to the propagation direction z is determined by the radial distribution functions, $\mathbf{v}_n^\nu(r)$ and $\mathbf{T}_n^\nu(r)$, respectively. They are, in general, complex functions, and are referred to as the mode shapes of the waveguide mode.

B. Normal mode theory

The main tool for analyzing scattering problems is the normal mode theory. It was laid out for layered waveguides by Auld,[12] and a brief review of the fundamental framework is given here. In normal mode theory, one makes use of the fact that an arbitrary field distribution can be expanded into so-called normal modes. These normal modes have to fulfill two main conditions. First, the set of modes has to be *complete*, which means that they are sufficient, by superposition, to describe any arbitrary field distribution. Questions of this kind have been treated by Kirrmann[25] for Lamb modes. In this paper, completeness of the waveguide modes is assumed. Second, the modes must be *orthogonal*, which means that the expansion of an arbitrary field into normal modes is unique. Mode orthogonality for layered plate waveguides was treated by Auld,[12] and was generalized to cylindrical waveguides by Ditri.[17] To be strict, the orthogonality relationships derived in these references are actually *biorthogonality* relationships, since they involve two different fields, namely the particle velocity and stress fields. We now review and consider these relationships.

For any two solutions, "1" and "2," of the particle velocity and stress fields to the field equations, a real reciprocity relationship of the following form holds:[12]

$$\boldsymbol{\nabla}\cdot\{\mathbf{v}_1\cdot\mathbf{T}_2-\mathbf{v}_2\cdot\mathbf{T}_1\}=0, \tag{3}$$

where the indices denote the solutions "1" and "2," respectively, and $\boldsymbol{\nabla}$ is the divergence operator. Within a volume V, where the material parameters are the same for solutions "1" and "2," one can therefore write for a given frequency ω,

$$\int_V \boldsymbol{\nabla}\cdot\{\mathbf{v}_1\cdot\mathbf{T}_2-\mathbf{v}_2\cdot\mathbf{T}_1\}dV=0. \tag{4}$$

Consider the case where solution "1" is given by mode (ν,n), and solution "2" is given by mode (μ,m). Equation (4) then becomes

$$\int_V \boldsymbol{\nabla}\cdot\{\mathbf{v}_n^\nu\cdot\mathbf{T}_m^\mu-\mathbf{v}_m^\mu\cdot\mathbf{T}_n^\nu\}dV=0. \tag{5}$$

Using the two-dimensional Gauss theorem, it follows that

$$\int_S \frac{\partial}{\partial z}\{\mathbf{v}_n^\nu\cdot\mathbf{T}_m^\mu-\mathbf{v}_m^\mu\cdot\mathbf{T}_n^\nu\}\hat{\mathbf{z}}\,dS+\oint_C\{\mathbf{v}_n^\nu\cdot\mathbf{T}_m^\mu-\mathbf{v}_m^\mu\cdot\mathbf{T}_n^\nu\}\hat{\mathbf{r}}\,dl=0, \tag{6}$$

where S is the cross section of the waveguide and C is a line along the circumference of S, $\hat{\mathbf{z}}$ is a unit vector in the propagation direction, and $\hat{\mathbf{r}}$ is a unit vector in the outward radial direction (see Fig. 1). For a free waveguide the tractions normal to the surface of the waveguide vanish, $\mathbf{T}\cdot\hat{\mathbf{r}}=0$, whereas for a waveguide with rigid boundaries the particle

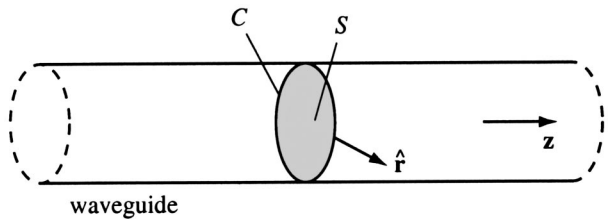


FIG. 1. Circular waveguide of cross section S and circumference C. $\hat{\mathbf{r}}$ is the outward normal vector to C.

velocities vanish, $\mathbf{v}=0$. Therefore, in both cases, Eq. (6) reduces to

$$\int_S \frac{\partial}{\partial z}\{\mathbf{v}_n^\nu \cdot \mathbf{T}_m^\mu - \mathbf{v}_m^\mu \cdot \mathbf{T}_n^\nu\}\hat{\mathbf{z}}\, dS = 0. \tag{7}$$

The dependence of the field variables on z implies that

$$-i(k_n + k_m)\int_S \{\mathbf{v}_n^\nu \cdot \mathbf{T}_m^\mu - \mathbf{v}_m^\mu \cdot \mathbf{T}_n^\nu\}\hat{\mathbf{z}}\, dS = 0. \tag{8}$$

From this equation, either $k_n = -k_m$, or the integral vanishes:

$$\int_S \{\mathbf{v}_n^\nu \cdot \mathbf{T}_m^\mu - \mathbf{v}_m^\mu \cdot \mathbf{T}_n^\nu\}\hat{\mathbf{z}}\, dS = 0, \quad \text{for } k_n \neq -k_m. \tag{9}$$

This is known as the real orthogonality relationship. However, there are cases where two different modes can have $k_n = -k_m$. In this case, orthogonality between the two modes is restored by a second relationship, called the angular orthogonality relationship. For $k_n = -k_m$, the integral on the left-hand side of Eq. (9) can be further evaluated (omitting the time dependence):

$$\int_S \{\mathbf{v}_n^\nu \cdot \mathbf{T}_m^\mu - \mathbf{v}_m^\mu \cdot \mathbf{T}_n^\nu\}\hat{\mathbf{z}}\, dS = \int_0^R \{\mathbf{v}_n^\nu \cdot \mathbf{T}_m^\mu - \mathbf{v}_m^\mu \cdot \mathbf{T}_n^\nu\}(r)\hat{\mathbf{z}} r\, dr \int_{\Theta_0}^{\Theta_0 + 2\pi} e^{i(\nu+\mu)\theta}\, d\Theta. \tag{10}$$

The integration over Θ shows that

$$\int_S \{\mathbf{v}_n^\nu \cdot \mathbf{T}_m^\mu - \mathbf{v}_m^\mu \cdot \mathbf{T}_n^\nu\}\hat{\mathbf{z}}\, dS = 0, \quad \text{for } \nu \neq -\mu. \tag{11}$$

This is called the angular orthogonality relationship. It is slightly different from the one presented by Ditri,[17] which, written in a form compatible with the mode naming here, states that

$$\int_{\Theta_0}^{\Theta_0 + 2\pi} \mathbf{v}_n^\nu \cdot \mathbf{T}_m^\mu\, d\Theta = 0, \quad \text{for } \upsilon \neq \mu. \tag{12}$$

This is due to the fact that the modal distribution fields in Ref. 17 are assumed to vary trigonometrically as either $\sin(\nu\Theta)$ or $\cos(\nu\Theta)$. However, in our case in which we think of the circumferential order as an angular wave number, the more general exponential notation, $e^{i\nu\Theta}$, is more consistent. The sine or cosine notation describes a standing wave solution along the circumference of the waveguide, and can be thought of as a superposition of two spiral waves of the exponential type, each in the opposite angular direction.

The orthogonality relationship for a cylinder in vacuum can thus be written in the following form:

$$\int_S \{\mathbf{v}_n^\nu \cdot \mathbf{T}_m^\mu - \mathbf{v}_m^\mu \cdot \mathbf{T}_n^\nu\}\hat{\mathbf{z}}\, dS = 0, \quad \text{for } k_n \neq -k_m \quad \text{or} \quad \nu \neq -\mu. \tag{13}$$

For the following scattering analysis it is convenient to introduce two more notations. First, we introduce the convention

$$-k_n = k_{-n}, \tag{14}$$

to label the same mode propagating in the negative z direction. We refer to modes whose energy propagates in the positive axial direction as ''forward'' modes, and those whose energy propagates in the negative axial direction as ''backward'' modes. Nonpropagating and inhomogeneous modes, however, do not propagate energy in any direction. One can imagine these modes as local vibrations, which exist only at waveguide discontinuities such as the edge of a plate or the surface discontinuity considered in this paper. In this case, the equivalent to a forward mode is a nonpropagating or inhomogeneous mode that is attenuated in the positive z direction.

Second, it is convenient to define the quantity $Q_{mn}^{\mu\nu}$ as

$$Q_{mn}^{\mu\nu} = -\frac{1}{4}\int_S \{\mathbf{v}_m^\mu \cdot \mathbf{T}_n^\nu - \mathbf{v}_n^\nu \cdot \mathbf{T}_m^\mu\}\hat{\mathbf{z}}\, dS. \tag{15}$$

For propagating modes, that means for modes with a purely real wave number, the fields satisfy the following relations (Ref. 12, p. 188):

$$\mathbf{v}_{-m}^{-\mu}(r,\Theta) = -\mathbf{v}_m^{\mu *}(r,\Theta), \tag{16}$$

$$\mathbf{T}_{-m}^{-\mu}(r,\Theta) = \mathbf{T}_m^{\mu *}(r,\Theta). \tag{17}$$

Evaluating $Q_{m-m}^{\mu-\mu}$ with these relationships yields

$$Q_{m-m}^{\mu-\mu} = -\frac{1}{4}\int_S \{\mathbf{v}_m^\mu \cdot \mathbf{T}_{-m}^{-\mu} - \mathbf{v}_{-m}^{-\mu} \cdot \mathbf{T}_m^\mu\}\hat{\mathbf{z}}\, dS \tag{18}$$

$$= -\frac{1}{4}\int_S \{\mathbf{v}_m^\mu \cdot \mathbf{T}_m^{\mu *} + \mathbf{v}_m^{\mu *} \cdot \mathbf{T}_m^\mu\}\hat{\mathbf{z}}\, dS = P_{mm}^{\mu\mu}, \tag{19}$$

where $P_{mm}^{\mu\mu}$ is the average power flow of mode (μ,m) propagating in the z direction. The average power flow is the real part of the complex power flow.

In scattering analysis, the modal fields are usually normalized such that the average power flow through the cross section of the waveguide is unity. However, nonpropagating and inhomogeneous modes do not carry power. Therefore they cannot be normalized this way. Some researchers used the complex power flow instead, noting that nonpropagating modes can be normalized to the imaginary unit.[19] However, in order to be able to apply the same normalization procedure to all modes, the modal fields will be normalized in this paper such that

$$Q_{m-m}^{\mu-\mu} = 1. \tag{20}$$

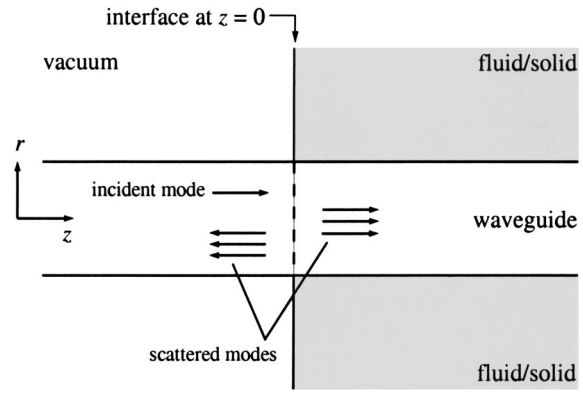


FIG. 2. A schematic of the problem considered in this paper. A propagating mode incident from the free section of the bar is scattered at the interface at $z=0$ due a change in surface impedance.

Moreover, propagating modes with purely real wave numbers are still power normalized by this expression, since

$$Q_{m-m}^{\mu-\mu}=P_{mm}^{\mu\mu}=1. \tag{21}$$

III. PROBLEM STATEMENT AND SOLUTION

An infinitely long cylindrical waveguide is half free and half embedded in another material, and the embedding material is assumed to extend infinitely in the radial direction (see Fig. 2). A propagating waveguide mode is incident from the free section of the waveguide and will be scattered and mode converted at the interface between the free and the embedded section due to the change in surface impedance. Two different approaches to solving the problem will be taken in the following. In Sec. III A, we follow the S-parameter formalism laid out by Auld.[12] This analysis yields an exact scattering formula for a guided wave incident from the free section of the cylinder. In Sec. III B it is shown that the same expression can be obtained by simply matching the fields over the cross section of the waveguide, thus neglecting the free surface of the embedding material at the interface between the free and the embedded section. Finally, in Sec. III C, a complete modal solution is given using a more intuitive notation.

A. S-parameter formalism

The first derivation follows closely the S-parameter method presented by Auld for the calculation of scattering coefficients.[12] Consider a waveguide with and without surface loading, as depicted in Fig. 3. At any point within the volume V, the reciprocity relation holds:

$$\nabla\cdot\{\mathbf{v}_1\cdot\mathbf{T}_2-\mathbf{v}_2\cdot\mathbf{T}_1\}=\nabla\cdot\{\cdots\}=0, \tag{22}$$

where the expression $\{\mathbf{v}_1\cdot\mathbf{T}_2-\mathbf{v}_2\cdot\mathbf{T}_1\}$ has been replaced by $\{\cdots\}$ for brevity. Therefore

$$\int_V \nabla\cdot\{\cdots\}dV=0. \tag{23}$$

Let the fields denoted by "1" be the solution for the case when there is no surface loading, and solution "2" for the case with surface loading. This is valid, because the reciprocity relation holds anywhere in the waveguide, where the material parameters are the same. The surface S of the volume V can be divided into three regions (see Fig. 3):

$$S=S_l+S_r+S_c, \tag{24}$$

where S_l is the plane on the left-hand side, S_r the plane on the right-hand side, and S_c the surface on the circumference of the cylindrical waveguide. By applying the Gauss theorem, it follows that

$$-\int_{S_l}\{\cdots\}\hat{\mathbf{z}}\,dS+\int_{S_r}\{\cdots\}\hat{\mathbf{z}}\,dS+\int_{S_c}\{\cdots\}\hat{\mathbf{r}}\,dS=0. \tag{25}$$

Consider a mode $(\overline{\mu},\overline{m})$ incident from the left, indicated as such with a bar superscript for clarity. For solution "1," no scattering occurs, and the fields on S_l and S_r can be written in the following form:

$$\mathbf{v}_1=\begin{cases}\mathbf{v}_{\overline{m}}^{\overline{\mu}}, & \text{on } S_l,\\ S_{\overline{mm}}^{\overline{\mu\mu}}\mathbf{v}_{\overline{m}}^{\overline{\mu}}, & \text{on } S_r;\end{cases} \tag{26}$$

$$\mathbf{T}_1=\begin{cases}\mathbf{T}_{\overline{m}}^{\overline{\mu}}, & \text{on } S_l,\\ S_{\overline{mm}}^{\overline{\mu\mu}}\mathbf{T}_{\overline{m}}^{\overline{\mu}}, & \text{on } S_r.\end{cases} \tag{27}$$

Here, the scattering coefficients, S, called the S parameters, are defined in a slightly different way than in Ref. 12. For example, $S_{nm}^{\nu\mu}$ refers to the coefficient for mode (μ,m) being incident from the left and mode (ν,n) being scattered (transmitted) to the right-hand side of the discontinuity, whereas $S_{-nm}^{-\nu\mu}$ denotes the reflection of $(-\nu,-n)$ to the left-hand side for an incident mode (μ,m). The mode propagation does not depend on z except for attenuation, whose effect is implicitly contained in the expressions for the particle velocity and the stress. Therefore it follows that

$$S_{\overline{mm}}^{\overline{\mu\mu}}=1. \tag{28}$$

Solution "2" must take the reflected and transmitted modes into account. For reasons that will become clear later, mode $(\overline{\nu},\overline{n})$ is chosen to be incident from the left. The fields can be expressed as

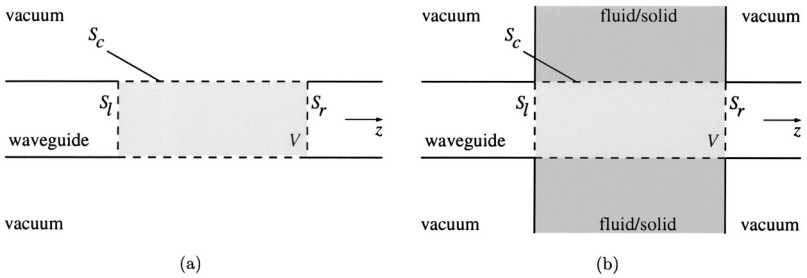


FIG. 3. (a) Solution "1," without surface loading, and (b) solution "2," with surface loading.

$$\mathbf{v}_2 = \begin{cases} \mathbf{v}_{\bar{n}}^{\bar{\nu}} + \sum_{n,\nu} S_{-nn}^{\nu\bar{\nu}} \mathbf{v}_{-n}^{\nu}, & \text{on } S_l, \\ \sum_{n,\nu} S_{nn}^{\nu\bar{\nu}} \mathbf{v}_n^{\nu}, & \text{on } S_r; \end{cases} \tag{29}$$

$$\mathbf{T}_2 = \begin{cases} \mathbf{T}_{\bar{n}}^{\bar{\nu}} + \sum_{n,\nu} S_{-nn}^{\nu\bar{\nu}} \mathbf{T}_{-n}^{\nu}, & \text{on } S_l, \\ \sum_{n,\nu} S_{nn}^{\nu\bar{\nu}} \mathbf{T}_n^{\nu}, & \text{on } S_r. \end{cases} \tag{30}$$

The sums run over negative and positive angular wave numbers, but over positive integers n only. Evaluating the integral from Eq. (25) on S_r yields

$$\int_{S_r} \{\cdots\}\hat{\mathbf{z}}\, dS = \int_{S_r} \sum_{n,\nu} S_{nn}^{\nu\bar{\nu}} \{\mathbf{v}_m^{\bar{\mu}} \cdot \mathbf{T}_n^{\nu} - \mathbf{v}_n^{\nu} \cdot \mathbf{T}_m^{\bar{\mu}}\}\hat{\mathbf{z}}\, dS, \tag{31}$$

$$= 0. \tag{32}$$

Note at this stage that therefore Eq. (25) becomes

$$\int_{S_l} \{\mathbf{v}_m^{\mu} \cdot \mathbf{T}_2 - \mathbf{v}_2 \cdot \mathbf{T}_m^{\mu}\}\hat{\mathbf{z}}\, dS = \int_{S_c} \{\mathbf{v}_m^{\mu} \cdot \mathbf{T}_2 - \mathbf{v}_2 \cdot \mathbf{T}_m^{\mu}\}\hat{\mathbf{r}}\, dS. \tag{33}$$

Further evaluating the integral on S_l, one obtains

$$\begin{aligned} -\int_{S_l} \{\cdots\}\hat{\mathbf{z}}\, dS &= -\int_{S_l} \{\mathbf{v}_m^{\bar{\mu}} \cdot \mathbf{T}_{\bar{n}}^{\bar{\nu}} - \mathbf{v}_{\bar{n}}^{\bar{\nu}} \cdot \mathbf{T}_m^{\bar{\mu}}\}\hat{\mathbf{z}}\, dS \\ &\quad - \int_{S_l} \sum_{n,\nu} S_{-nn}^{\nu\bar{\nu}} \{\mathbf{v}_m^{\bar{\mu}} \cdot \mathbf{T}_{-n}^{\nu} - \mathbf{v}_{-n}^{\nu} \cdot \mathbf{T}_m^{\bar{\mu}}\}\hat{\mathbf{z}}\, dS \\ &= 4S_{-mn}^{-\bar{\mu}\bar{\nu}}. \end{aligned} \tag{34}$$

Equation (25) therefore becomes

$$4S_{-mn}^{-\bar{\mu}\bar{\nu}} + \int_{S_c} \{\mathbf{v}_m^{\bar{\mu}} \cdot \mathbf{T}_2 - \mathbf{v}_2 \cdot \mathbf{T}_m^{\bar{\mu}}\}\hat{\mathbf{r}}\, dS = 0. \tag{35}$$

This is valid for an arbitrary mode $(\bar{\mu},\bar{m})$, which can be renamed (μ,m). Thus

$$S_{-mn}^{-\mu\bar{\nu}} = -\frac{1}{4}\int_{S_c} \{\mathbf{v}_m^{\mu} \cdot \mathbf{T}_2 - \mathbf{v}_2 \cdot \mathbf{T}_m^{\mu}\}\hat{\mathbf{r}}\, dS. \tag{36}$$

This is a general scattering formula for an incident mode $(\bar{\nu},\bar{n})$ being scattered into mode $(-\mu,-m)$. It is equivalent to the general scattering formula presented by Auld. This is an exact formula, and the S parameters can be calculated if the fields in "2" are known. It can be seen that the scattering coefficients will depend on the total waveguide surface area covered by the embedding material. To obtain the case considered here, the integral over the surface S_c is extended to $z=\infty$.

B. Alternative calculation

It will be shown in this section that the same result can be obtained by simply matching the boundary conditions at the interface where the waveguide enters the embedding material. Only the mode shapes over the adjoining cross section have to be considered in order to obtain an exact solution for modes that are incident from the free section. Thus, the free surface of the embedding material at $z=0$ does not have to be considered.

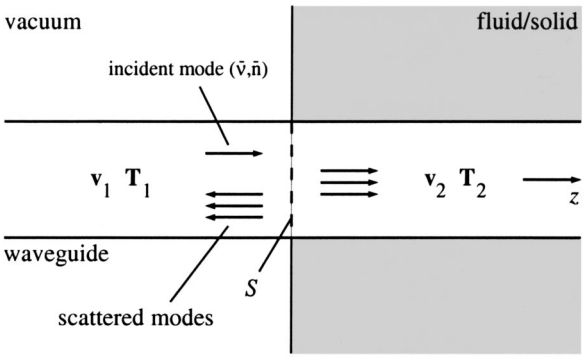


FIG. 4. A partly embedded circular waveguide. An incident mode $(\bar{\nu},\bar{n})$ is scattered at the interface between the free and embedded sections of the waveguide. The fields $\mathbf{v}_1$ and $\mathbf{T}_1$ in the free section are represented by a superposition of the incident and reflected waveguide modes. The fields excited in the embedded section are denoted by $\mathbf{v}_2$ and $\mathbf{T}_2$.

It is assumed that the field in the waveguide on the left-hand side of the interface between the vacuum and embedded parts, "1," is a superposition of the waveguide modes in the free cylinder (see Fig. 4); we refer to these here as the "vacuum" modes. The particle velocity and stress fields "1" are expressed in terms of the incident vacuum mode $(\bar{\nu},\bar{n})$, propagating in the $+z$ direction, and the sum of all reflected vacuum modes, propagating in the $-z$ direction. The sums run over negative and positive angular wave numbers, but over positive axial wave numbers only:

$$\mathbf{v}_1 = \mathbf{v}_{\bar{n}}^{\bar{\nu}} + \sum_{\nu,n} S_{-nn}^{\nu\bar{\nu}} \mathbf{v}_{-n}^{\nu}, \tag{37}$$

$$\mathbf{T}_1 = \mathbf{T}_{\bar{n}}^{\bar{\nu}} + \sum_{\nu,n} S_{-nn}^{\nu\bar{\nu}} \mathbf{T}_{-n}^{\nu}. \tag{38}$$

We will make assumptions about the fields in "2" later in this paper, but for the moment the fields created in the embedded section by the incident mode will just be denoted by $\mathbf{v}_2$ and $\mathbf{T}_2$. The boundary conditions within the waveguide at the interface between the free and the embedded section must be fulfilled:

$$\mathbf{v}_1 = \mathbf{v}_2, \tag{39}$$

$$\mathbf{T}_1 \cdot \hat{\mathbf{z}} = \mathbf{T}_2 \cdot \hat{\mathbf{z}}, \tag{40}$$

where $\hat{\mathbf{z}}$ is the unit vector in the z direction. The boundary conditions therefore become

$$\mathbf{v}_{\bar{n}}^{\bar{\nu}} + \sum_{\nu,n} S_{-nn}^{\nu\bar{\nu}} \mathbf{v}_{-n}^{\nu} = \mathbf{v}_2, \tag{41}$$

$$\mathbf{T}_{\bar{n}}^{\bar{\nu}} + \sum_{\nu,n} S_{-nn}^{\nu\bar{\nu}} \mathbf{T}_{-n}^{\nu} = \mathbf{T}_2. \tag{42}$$

Multiplying the first equation by $\mathbf{T}_m^{\mu}$ and the second equation by $\mathbf{v}_m^{\mu}$, one obtains

$$\mathbf{v}_{\bar{n}}^{\bar{\nu}} \cdot \mathbf{T}_m^{\mu} + \sum_{\nu,n} S_{-nn}^{\nu\bar{\nu}} \mathbf{v}_{-n}^{\nu} \cdot \mathbf{T}_m^{\mu} = \mathbf{v}_2 \cdot \mathbf{T}_m^{\mu}, \tag{43}$$

$$\mathbf{v}_m^{\mu}\cdot\mathbf{T}_{\bar{n}}^{\bar{\nu}}+\sum_{\nu,n} S_{-n,\bar{n}}^{\nu,\bar{\nu}}\mathbf{v}_m^{\mu}\cdot\mathbf{T}_{-n}^{\nu}=\mathbf{v}_m^{\mu}\cdot\mathbf{T}_2. \tag{44}$$

Subtracting Eq. (44) from (43) yields

$$\{\mathbf{v}_{\bar{n}}^{\bar{\nu}}\cdot\mathbf{T}_m^{\mu}-\mathbf{v}_m^{\mu}\cdot\mathbf{T}_{\bar{n}}^{\bar{\nu}}\}+\sum_{\nu,n} S_{-n\bar{n}}^{\nu\bar{\nu}}\{\mathbf{v}_{-n}^{\nu}\cdot\mathbf{T}_m^{\mu}-\mathbf{v}_m^{\mu}\cdot\mathbf{T}_{-n}^{\nu}\}$$
$$=\{\mathbf{v}_2\cdot\mathbf{T}_m^{\mu}-\mathbf{v}_m^{\mu}\cdot\mathbf{T}_2\}. \tag{45}$$

In order to take advantage of the real orthogonality relation, this is integrated over the cross section S. The terms on the left-hand side can be examined separately:

$$\int_S\{\mathbf{v}_{\bar{n}}^{\bar{\nu}}\cdot\mathbf{T}_m^{\mu}-\mathbf{v}_m^{\mu}\cdot\mathbf{T}_{\bar{n}}^{\bar{\nu}}\}\hat{\mathbf{z}}\,dS=0, \tag{46}$$

$$\int_S\sum_{\nu,n} S_{-n\bar{n}}^{\nu\bar{\nu}}\{\mathbf{v}_{-n}^{\nu}\cdot\mathbf{T}_m^{\mu}-\mathbf{v}_m^{\mu}\cdot\mathbf{T}_{-n}^{\nu}\}\hat{\mathbf{z}}\,dS=4S_{-m\bar{n}}^{-\mu\bar{\nu}}. \tag{47}$$

Replacing this into Eq. (45), the S parameter for the scattered mode $(-\mu,-m)$ and the incident mode $(\bar{\nu},\bar{n})$ is obtained:

$$S_{-m\bar{n}}^{-\mu\bar{\nu}}=-\frac{1}{4}\int_S\{\mathbf{v}_m^{\mu}\cdot\mathbf{T}_2-\mathbf{v}_2\cdot\mathbf{T}_m^{\mu}\}\hat{\mathbf{z}}\,dS. \tag{48}$$

The surface S is equivalent to S_l in Eq. (33). Since the boundary conditions are fulfilled on this surface, the fields $\mathbf{v}_2$ and $\mathbf{T}_2$ on S in the two approaches to calculate the scattering coefficients are the same. Thus, it follows that the expressions in Eqs. (36) and (48) are equivalent.

Similarly, by multiplying Eqs. (41) and (42) by $\mathbf{T}_{-m}^{\mu}$ and $\mathbf{v}_{-m}^{\mu}$, respectively, it can be shown that

$$-\frac{1}{4}\int_S\{\mathbf{v}_2\cdot\mathbf{T}_{-m}^{\mu}-\mathbf{v}_{-m}^{\mu}\cdot\mathbf{T}_2\}\hat{\mathbf{z}}\,dS$$
$$=\begin{cases}1, & \text{for } \mu=-\bar{\nu} \text{ and } m=\bar{n},\\ 0, & \text{else.}\end{cases} \tag{49}$$

Thus, the scattering parameters for the case of a mode incident from the free part of the waveguide can be calculated by just considering the integrals over the cross section of the waveguide, ignoring the free surface of the embedding material. However, this is not necessarily true for the case when a mode is incident from the embedded section of the waveguide. First of all, orthogonality would have to be shown for these modes. Whether or not it is sufficient only to consider the adjoining cross section in the embedded part would then depend on the nature of this orthogonality relationship.

C. Modal solution

In fact, Eqs. (48) and (49) already give the complete solution to the problem in the case where modes are incident from the free section of the waveguide. However, it cannot readily be seen how to calculate the scattering parameters from this. Therefore in this section the complete modal solution for the case of modes that are incident from the free section of the waveguide will be derived in a more intuitive notation similar to that presented by Engan.[19]

The analysis in this section is based on the assumption that both the vacuum and the embedded modes form a complete set of functions. Orthogonality of the vacuum modes has been shown in Sec. II B, and will be used in the derivations. While orthogonality for the embedded modes is in this case not necessary, it is, however, preferable in order to obtain a unique solution to the scattering problem. Orthogonality of embedded waveguide modes is discussed in more detail in Sec. IV.

We will restrict the following derivation to longitudinal modes, which are of circumferential order zero. The superscripts denoting the circumferential order will be omitted from now on, since by symmetry order zero modes can only scatter into modes of the same order, and also torsional modes cannot scatter into longitudinal modes, and *vice versa*.

If the amplitudes of incoming modes are denoted by a's, and scattered mode amplitudes are denoted by b's, the particle velocity and stress fields in the free section of the bar can be written as

$$\mathbf{v}_1=\sum_m a_m\mathbf{v}_m+b_{-m}\mathbf{v}_{-m}, \tag{50}$$

$$\mathbf{T}_1=\sum_m a_m\mathbf{T}_m+b_{-m}\mathbf{T}_{-m}. \tag{51}$$

The index of the mode amplitudes indicate in which direction the scattered mode is propagating. As stated earlier, for nonpropagating and inhomogeneous modes, this is chosen to be the direction of the decay. It is now assumed that the fields being excited in "2" are composed of the modes of an embedded cylinder, denoted by a dash:

$$\mathbf{v}_2=\sum_{m'} b_{m'}\mathbf{v}_{m'}, \tag{52}$$

$$\mathbf{T}_2=\sum_{m'} b_{m'}\mathbf{T}_{m'}. \tag{53}$$

Note that there is no incident mode from the embedded section. In principle, the sums in Eqs. (50)–(53) extend over an infinite number of modes. However, for the numerical calculations, these sums have to be truncated after a certain mode M.

One can collect the incoming and scattered mode amplitudes in column vectors, and they are related to each other by a scattering matrix in the following way:

$$\begin{bmatrix} b_{-1}\\ b_{-2}\\ \vdots\\ b_{-M}\\ b_{1'}\\ b_{2'}\\ \vdots\\ b_{M'}\end{bmatrix}=\begin{bmatrix} S_{-11} & S_{-12} & \cdots & S_{-1M}\\ S_{-21} & S_{-22} & \cdots & S_{-2M}\\ \vdots & \vdots & \ddots & \vdots\\ S_{-M1} & S_{-M2} & \cdots & S_{-MM}\\ S_{1'1} & S_{1'2} & \cdots & S_{1'M}\\ S_{2'1} & S_{2'2} & \cdots & S_{2'M}\\ \vdots & \vdots & \ddots & \vdots\\ S_{M'1} & S_{M'2} & \cdots & S_{M'M}\end{bmatrix}\begin{bmatrix} a_1\\ a_2\\ \vdots\\ a_M\end{bmatrix}$$
$$=\begin{bmatrix}[S_{-mn}]\\ [S_{m'n}]\end{bmatrix}\begin{bmatrix} a_1\\ a_2\\ \vdots\\ a_M\end{bmatrix}. \tag{54}$$

The scattering submatrices S_{-mn} and $S_{m'n}$ are $M\times M$ matrices, whereas the full scattering matrix is an $M\times 2M$ matrix. First, the boundary conditions are applied on the interface between the free and the embedded section [see Eqs. (39) and (40), respectively]:

$$\sum_m a_m\mathbf{v}_m + b_{-m}\mathbf{v}_{-m} = \sum_{m'} b_{m'}\mathbf{v}_{m'}, \tag{55}$$

$$\sum_m a_m\mathbf{T}_m + b_{-m}\mathbf{T}_{-m} = \sum_{m'} b_{m'}\mathbf{T}_{m'}. \tag{56}$$

Note that it has been shown in Secs. III A and III B that it is only necessary to invoke these boundary conditions over the cross section of the waveguide in order to obtain an exact solution. Multiplying Eq. (55) by $\mathbf{T}_{-n}$ and Eq. (56) by $\mathbf{v}_{-n}$, and subtracting (56) from (55) yields

$$\sum_m a_m(\mathbf{v}_m\cdot\mathbf{T}_{-n} - \mathbf{v}_{-n}\cdot\mathbf{T}_m) + \sum_m b_{-m}(\mathbf{v}_{-m}\cdot\mathbf{T}_{-n} - \mathbf{v}_{-n}\cdot\mathbf{T}_{-m})$$

$$= \sum_{m'} b_{m'}(\mathbf{v}_{m'}\cdot\mathbf{T}_{-n} - \mathbf{v}_{-n}\cdot\mathbf{T}_{m'}). \tag{57}$$

Then, by integrating over the cross section of the waveguide, one makes use of the orthogonality relation (13), and some terms in Eq. (57) vanish. Noting that the fields are normalized according to Eq. (20), it follows that

$$a_n = -\frac{1}{4}\sum_{m'} b_{m'}\int_S (\mathbf{v}_{m'}\cdot\mathbf{T}_{-n} - \mathbf{v}_{-n}\cdot\mathbf{T}_{m'})\cdot\hat{\mathbf{z}}\,dS$$

$$= -\sum_{m'} b_{m'}Q_{-nm'}. \tag{58}$$

This can be written in matrix form:

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix} = -\begin{bmatrix} Q_{-11'} & Q_{-12'} & \cdots & Q_{-1M'} \\ Q_{-21'} & Q_{-22'} & \cdots & Q_{-2M'} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{-M1'} & Q_{-M2'} & \cdots & Q_{-MM'} \end{bmatrix}\begin{bmatrix} b_{1'} \\ b_{2'} \\ \vdots \\ b_{M'} \end{bmatrix}. \tag{59}$$

In an abbreviated notation, this is

$$[a_n] = -[Q_{-nm'}][b_{m'}], \tag{60}$$

so

$$[b_{m'}] = -[Q_{-nm'}]^{-1}[a_n]. \tag{61}$$

Second, multiplying Eq. (55) by $\mathbf{T}_n$ and Eq. (56) by $\mathbf{v}_n$, and subtracting (56) from (55) yields

$$\sum_m a_m(\mathbf{v}_m\cdot\mathbf{T}_n - \mathbf{v}_n\cdot\mathbf{T}_m) + \sum_m b_{-m}(\mathbf{v}_{-m}\cdot\mathbf{T}_n - \mathbf{v}_n\cdot\mathbf{T}_{-m})$$

$$= \sum_{m'} b_{m'}(\mathbf{v}_{m'}\cdot\mathbf{T}_n - \mathbf{v}_n\cdot\mathbf{T}_{m'}). \tag{62}$$

Again, this is integrated over the cross section of the waveguide, and some terms vanish due to orthogonality. One is left with the following equation:

$$b_{-n} = \frac{1}{4}\sum_{m'} b_{m'}\int_S (\mathbf{v}_{m'}\cdot\mathbf{T}_n - \mathbf{v}_n\cdot\mathbf{T}_{m'})\cdot\hat{\mathbf{z}}\,dS$$

$$= \sum_{m'} b_{m'}Q_{nm'}. \tag{63}$$

Replacing n by m, this can be abbreviated to give

$$[b_{-m}] = [Q_{mm'}][b_{m'}]. \tag{64}$$

Recalling Eq. (61), the coefficients b_{-m} are calculated as follows:

$$[b_{-m}] = -[Q_{mm'}][Q_{-nm'}]^{-1}[a_n]. \tag{65}$$

Finally, Eqs. (61) and (65) can be combined to give the solution to the scattering problem:

$$\begin{bmatrix} [b_{-m}] \\ [b_{m'}] \end{bmatrix} = \begin{bmatrix} -[Q_{mm'}][Q_{-nm'}]^{-1} \\ -[Q_{-nm'}]^{-1} \end{bmatrix}[a_n]. \tag{66}$$

By comparison with the general scattering formula in Eq. (54), it follows that

$$[S_{-mn}] = -[Q_{mm'}][Q_{-nm'}]^{-1} \tag{67}$$

and

$$[S_{m'n}] = -[Q_{-nm'}]^{-1}. \tag{68}$$

The scattering amplitudes are calculated here using a normalization procedure for the modal fields according to Eq. (20). For the interpretation of the results, it is convenient to state the scattering amplitudes in terms of the scattered power. The power normalized scattering coefficients, denoted by capital letters A_m, B_{-m}, and $B_{m'}$, respectively, are obtained from the amplitudes calculated here, a_m, b_{-m}, and $b_{m'}$, respectively, by multiplying by the square root of the power of the corresponding mode, for example:

$$B'_m = b'_m\sqrt{P_{m'm'}}. \tag{69}$$

Note that for nonpropagating and inhomogeneous modes, $P_{mm}=0$, and therefore no power is carried away from the scattering region by these modes.

IV. NUMERICAL IMPLEMENTATION

A. Scattering matrix calculation

The calculation of the elements of the scattering matrix requires the knowledge of the radial distribution functions of the modes that are included in the solution. In most cases, a closed form analytical solution of the modal fields cannot be obtained, and the dispersion equation has to be solved numerically. In this study, the general purpose software tool DISPERSE[3] has been used to trace the dispersion curves and to obtain the particle velocity and stress mode shapes of each mode. The dispersion curves of both the vacuum and the embedded modes are found using a global matrix approach for multilayered waveguide structures.[26] The dispersion equation for a particular system can be expressed in the form $f(k_{\mathrm{re}},k_{\mathrm{im}},\omega)=0$, with the roots of f occurring for particular combinations of real wave number, imaginary wave number,

and frequency. Details regarding the theory for embedded cylindrical structures and the numerical root finding procedure can be found in Ref. 27. The mode shapes of the waveguide modes found in this way are then used to calculate the elements of the scattering matrix. Typically, a resolution of 200 points over the radius of the waveguide was found to give a reasonable accuracy for the calculation of the integrals.

In order to ensure that the submatrix $[Q_{m'-n}]$ in Eq. (60) is invertible, the equivalent number of modes in the embedded cylinder to that in the free cylinder will be included in the modal solution.

The energy balance is often used as an estimate of the error in the calculation of the scattering coefficients,[14] and it is adopted here. After normalization of the coefficients to represent the scattered power [see Eq. (69)], the error ϵ in the total power flow is given by

$$\sum_m (|A_m|^2 - |B_{-m}|^2) - \sum_{m'} |B_{m'}|^2 = \epsilon. \tag{70}$$

For an exact solution, $\epsilon = 0$. It should be noted that the sums only run over propagating modes, since nonpropagating and inhomogeneous modes do not contribute to the power flow. Also note that modes propagating in the $-z$ direction have negative power flow. In the case that only one mode of unit power flow is incident, this simplifies to

$$1 - \sum_m |B_{-m}|^2 - \sum_{m'} |B_{m'}|^2 = \epsilon. \tag{71}$$

B. Orthogonality considerations in embedded waveguides

A key assumption in the modal analysis is that the fields in the embedded section can be represented by a sum of waveguide modes of the infinitely long embedded cylinder. In order for this solution to be unique, orthogonality of the embedded waveguide modes is required. The embedded modes are, in general, leaky modes, and there exist proofs of orthogonality in the literature for leaky layered waveguides.[28,29] However, it has to be mentioned that it is usually assumed that the wave number in the propagation direction is either purely real or purely imaginary, or that the fields vanish far away from the interface. This is not necessarily true, as leaky guided waves have a complex axial wave number. If they have a purely real wave number, they are essentially nonleaky modes. Also, the assumption that the fields vanish far away from the waveguide is not generally valid. This is illustrated by the example of a leaky wave in an embedded plate waveguide:

Figure 5 shows a schematic of a plate waveguide of thickness d, which is embedded on both sides in another material. The embedding material extends infinitely in the positive and negative y directions, and the waveguide and the embedding medium are assumed to be perfectly elastic. The waves that radiate into the embedding half-spaces are then plane waves that have a constant amplitude along their direction of propagation. The angle, γ, between the direction of propagation of the radiating waves and the guided wave is

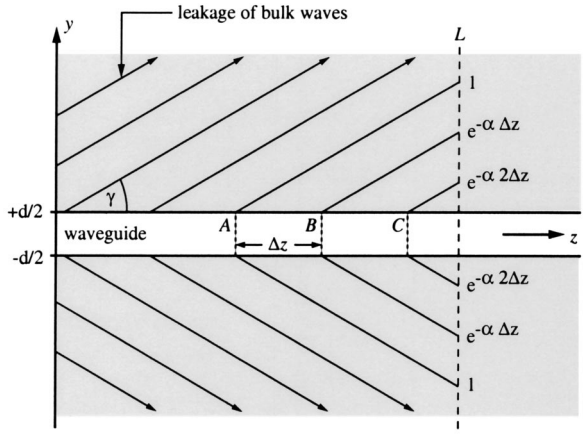


FIG. 5. A schematic showing bulk waves radiated from a plate waveguide of thickness d into the embedding material. See the text for a detailed discussion.

governed by Snell's law. Consider now the rays of plane waves radiating at three different positions (A, B, and C in Fig. 5) along the waveguide, each separated by a distance Δz. When leakage occurs, the guided wave that propagates in the z direction is attenuated. Let the amplitude of the plane wave at point A be unity, and the attenuation rate of the guided wave α. It follows then that the plane wave radiated at point B has an amplitude that is smaller by a factor of $e^{-\alpha \Delta z}$, at point C it is smaller by a factor of $e^{-\alpha 2 \Delta z}$, and so on. When the modal field is now considered along a line "L" in the y direction, it can be seen that the amplitudes at the intersection with the rays of the radiating waves increases with distance from the waveguide surface. To be more exact, one can consider Snell's law at the interface, which states that the real and the imaginary part of the wave numbers of the radiating waves and the guided wave in the z direction must be equal. It follows that the leaking plane waves must

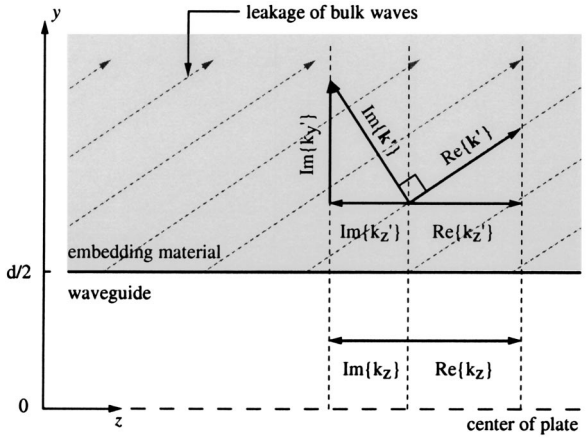


FIG. 6. A schematic showing the geometrical relationship between the complex wave number of the guided wave in the axial direction, k_z, and the complex wave number vector of the leaky inhomogeneous bulk wave, $\mathbf{k}'$. Note that $\mathrm{Im}\{k_z\} \leq 0$ corresponds to a guided wave that is attenuated in the positive z direction. In order to satisfy Snell's law and the condition that the imaginary part of the wave number of the leaky wave, $\mathrm{Im}\{\mathbf{k}'\}$, must be normal to its real part, $\mathrm{Re}\{\mathbf{k}'\}$ (see, for example, Ref. 30), there must be an imaginary wave number component, $\mathrm{Im}\{k_y'\}$, in the y direction away from the waveguide.

be, in fact, *inhomogeneous*, with an imaginary wave number component orthogonal to the real wave number component.[30] Geometric considerations show that the radiating inhomogeneous plane waves must have a wave number component in the *y* direction that causes them to increase exponentially in the *y* direction away from the waveguide surface (see Fig. 6). The fact that the fields tend to infinity away from the waveguide does not violate any radiation conditions since these fields originate from previously radiated waves. The situation is similar, though more complicated, for leaky waves in cylindrical systems since it involves complex Bessel functions. Therefore, integrals in an orthogonality relationship similar to Eq. (13), which include the half-spaces of the embedding medium, may be divergent. The question arises as to whether an orthogonality relationship for leaky waveguide modes has to be defined over the cross section of the waveguide or has to include the half-spaces. The solution to this problem is not strictly necessary in the scope of this investigation, however, it is an interesting subject for further research.

Orthogonality is assured for the waveguide modes of the free bar [see Eq. (13)]. This relationship was checked numerically using DISPERSE for the free bar modes in order to give an indication as to how accurate the numerical values are. It was found that, once the fields have been normalized according to Eq. (20), the value of $Q_{nm}^{\nu\mu}$ for two orthogonal modes was of the order of 10^{-5}. Strictly this should be zero, so this value represents the error in the calculations. For the equivalent modes of a cylinder embedded in an epoxy resin, as modeled in this paper, $Q_{n'm'}^{\nu'\mu'}$, with the integral defined over the cross section of the waveguide, was found to be typically of the order of 10^{-3}. Given the increased difficulty in finding the complex roots of the dispersion equation and the limited precision of the complex Bessel functions that have to be employed, this seems to be an acceptably small value. However, this numerical demonstration is not a proof, and it has to be accepted that orthogonality for leaky waves may have to be evaluated, including the embedding half-space and not only the waveguide cross section.

C. Dispersion curves

The scope of the following examples will be limited to the case where the lowest-order longitudinal mode, L(0,1), is incident from the free section of the cylinder. At relatively low frequencies, this has a practical application for the measurement of the shear properties of an embedding material.[1]

The dispersion curves can be presented in a three-dimensional representation, showing the real and imaginary part of the axial wave number, *k*, in the propagation direction against the frequency *f*. A summary of the material parameters used in the analysis is given in Table I.

TABLE I. Material parameters used for the analysis.

Material	Long. bulk velocity (m/s)	Shear bulk velocity (m/s)	Density (kg/m^3)
Steel	5960	3260	7932
Epoxy	2610	1000	1170
Material "a"	2000	1000	1000
Material "b"	4000	1000	1000

Onoe and Zemanek have presented the dispersion curves for a cylinder in vacuum in great detail, showing the real, imaginary as well as the complex branches.[16,31] The curves will therefore only be shown here for explanatory purposes. Real, imaginary, and complex branches correspond, respectively, to propagating, nonpropagating, and inhomogeneous solutions. As pointed out in Ref. 16, for every complex branch there exists an exact negative complex conjugate one. A three-dimensional representation of the first three longitudinal modes is given in Fig. 7(a). Consider, for example, the modes L(0,2) and L(0,3), which are both inhomogeneous between zero frequency and the cut-off frequency of L(0,2). In the inhomogeneous region, their imaginary wave numbers are identical, whereas the real parts of their wave number have equal magnitude and opposite sign. As an example, in order to show this relationship more clearly, the real wave number of the forward modes is plotted together with that of the backward modes in Fig. 7(b). The wave number of a forward mode has the same magnitude and opposite sign to the wave number of the corresponding backward mode, and therefore this graph allows a direct comparison of the real wave numbers of L(0,2) and L(0,3) in the inhomogeneous region by looking at L(0,2) and $-$L(0,3). The latter overlap completely in the inhomogeneous region, therefore the real wave number of L(0,2) and L(0,3) are exactly negative of each other.

The real part of the wave number is nonzero for an inhomogeneous mode and represents a phase propagation,

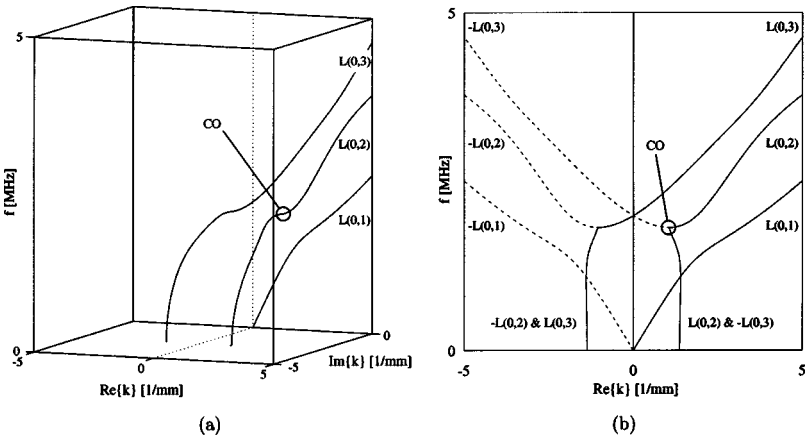


FIG. 7. (a) Three-dimensional dispersion curves for the first three longitudinal forward modes in a 1 mm steel cylinder in vacuum. (b) The real wave number of the first three longitudinal modes, showing forward and backward modes plotted in solid lines and dashed lines, respectively. CO indicates the cut-off point below which L(0,2) is inhomogeneous.

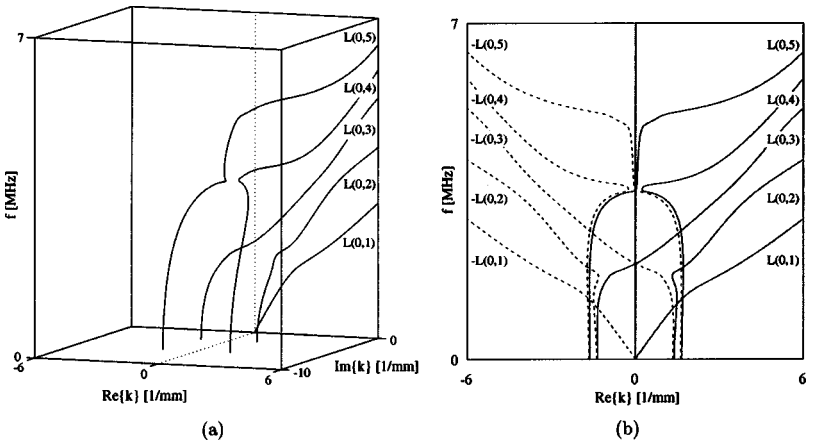


FIG. 8. (a) Three-dimensional dispersion curves for the first five longitudinal forward modes in a 1 mm steel cylinder embedded in epoxy resin. (b) The real wave number of the first five longitudinal modes, showing forward and also backward modes in solid lines and dashed lines, respectively.

whereas the imaginary part represents a spatial attenuation. However, as a simple power flow calculation shows, inhomogeneous modes do not carry power.

The situation is different for an embedded waveguide. Figure 8(a) shows the dispersion curves for the longitudinal modes L(0,1) to L(0,5) in a 1 mm steel cylinder embedded in an epoxy resin. In addition, the projection of these curves onto the real wave number–frequency plane is given in Fig. 8(b), which also shows the modes that propagate in the negative direction. From this it can clearly be seen that the modes that are equivalent to the inhomogeneous modes in the free cylinder do not have a negative complex conjugate counterpart. Instead, the modes seem to repel each other. Also, a power flow calculation shows that these modes do, in general, propagate power, and are therefore propagating modes.

The case where the embedding material is replaced by a perfectly rigid material is also discussed. It presents an appropriate validation case for two reasons. First, the validity of the modal solution can be tested without the assumption of orthogonality in the embedded case since the modes in a cylinder with rigid boundaries obey the orthogonality relationship presented in Eq. (7). Second, a rigid boundary can be thought of as a limiting case of infinite impedance of the embedding material. The dispersion curves are shown in Fig. 9. Note that there are only nonpropagating and inhomogeneous longitudinal modes in the low-frequency region and no propagating longitudinal modes. The consequences arising out of this are studied in Sec. VI.

V. VALIDATION BY FINITE ELEMENT MODELING

The Finite Element (FE) method is used here to validate the results obtained from the modal scattering analysis.

The FE analysis was carried out using the software FINEL, developed at Imperial College.[32] It performs efficient modeling of elastic wave propagation using a time-marching procedure. To model a circular waveguide, four-noded quadrilateral elements were used in axisymmetry. To obtain a stable solution, it is necessary to meet a number of requirements.[20] For numerical accuracy, the element size should at most be one-eighth of the shortest wavelength present. A rapidly varying mode shape over the cross section has to be accounted for correctly as well. The time step for the marching procedure should not exceed 0.8 times the time in which the fastest wave can propagate one element length.

The nature of the reflection of L(0,1) from the point where a steel cylinder enters a much larger diameter cylinder of epoxy resin was also examined in detail. A schematic of the FE model is shown in Fig. 10. The analysis covered the range from 0–0.5 MHz, and three different wire radii were modeled: 0.25, 0.5, and 0.75 mm. A one-cycle Hanning-windowed piston profile of 0.2 MHz center frequency was applied at the free end of the cylinder in order to obtain the reflection coefficient over a range of frequencies. Because the excitation has no significant energy above the cut-off frequency of L(0,2) for the chosen waveguide diameters, the only propagating mode excited by this configuration is

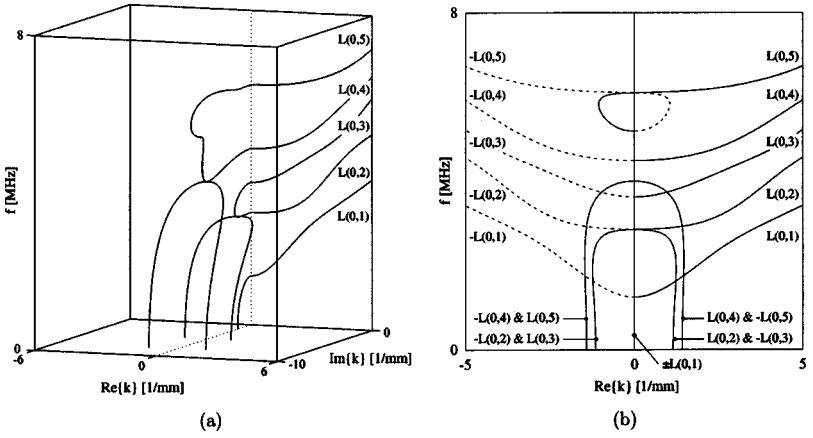


FIG. 9. (a) Three-dimensional dispersion curves for the first five longitudinal forward modes in a 1 mm steel cylinder with rigid boundaries. (b) The real wave number of the first five longitudinal modes, showing forward and also backward modes in solid lines and dashed lines, respectively. Note that there are no propagating modes at low frequencies.

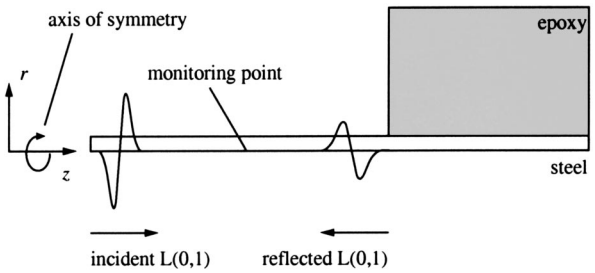


FIG. 10. A schematic of the Finite Element model for a cylinder entering an epoxy resin.

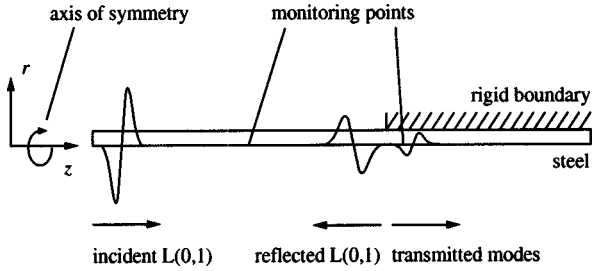


FIG. 12. A schematic of the Finite Element model for a cylinder entering a region with rigid boundaries.

L(0,1). One has to make sure that the reflection from the entry point is not disturbed by any other reflections from the end of the cylinder or from the boundaries of the epoxy by choosing a sufficiently long embedded length and a sufficiently large outer diameter of the epoxy. The monitoring point for displacements in the axial direction is halfway between the free end of the cylinder and the entry point.

A typical FE time trace is shown in Fig. 11. The first signal corresponds to the incident mode L(0,1) passing the monitoring point, whereas the second signal is the reflection from the entry point into the epoxy. Reflections from the other boundaries of the embedded part of the model come later and can thus be neglected. A Fourier transform was then applied to both signals individually. The division of the amplitude spectrum of the reflected signal by the amplitude spectrum of the incident signal gives the frequency-dependent amplitude reflection coefficient.

The reflection from the interface between a cylinder with a free boundary and one with a perfectly rigid boundary was also examined. A schematic of the FE model is given in Fig. 12. The radius of the steel cylinder in this analysis was 0.5 mm. Since there are no propagating modes in the cylinder with rigid boundaries in the frequency range covered by the analysis, only nonpropagating and inhomogeneous modes will be excited in the rigid region. In order to observe these modes, the displacements were additionally monitored in the axial direction in steps of 0.125 mm to the right of the interface on the axis of the cylinder.

VI. RESULTS AND DISCUSSION

A. Cylinder with rigid boundary

The reflection of L(0,1) from the interface between a cylinder with free and rigid boundaries is considered mainly for the validation of the modal solution, since orthogonality is given in the cylinder with a rigid surface. However, it also gives some insight into the behavior of nonpropagating and inhomogeneous modes.

As stated earlier, there are no propagating modes at low frequencies in the rigid cylinder. When the L(0,1) mode is incident from the free cylinder, then total reflection must occur at the interface at frequencies that are below the cut-off frequency of L(0,2) in the free and L(0,1) in the rigid cylinder. This is because the nonpropagating and inhomogeneous modes in both parts of the cylinder cannot propagate energy away from the scattering region. Therefore, both the modal and the FE solution should give a reflection coefficient of unity. As shown in Fig. 13, this is confirmed by both methods to a high degree of numerical accuracy. To obtain this result, only the propagating L(0,1) mode in the free section and the

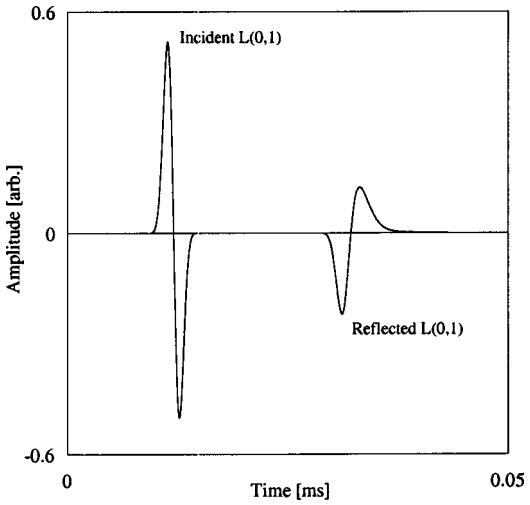


FIG. 11. A typical Finite Element time trace of the incident and reflected L(0,1) mode in a 0.5 mm diameter steel bar that is partly embedded in epoxy.

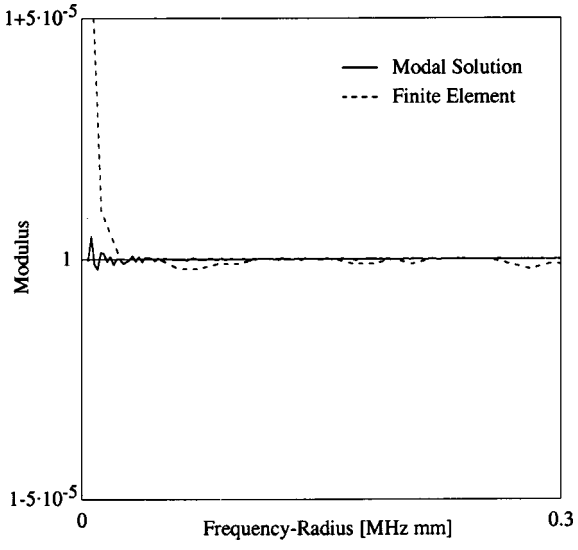


FIG. 13. The modulus of the reflection coefficient of L(0,1) from the interface between a cylinder with a free surface and one with a rigid surface. As explained in the text, this should be exactly unity. Note that the ordinate has been scaled to better show the error in the calculations.

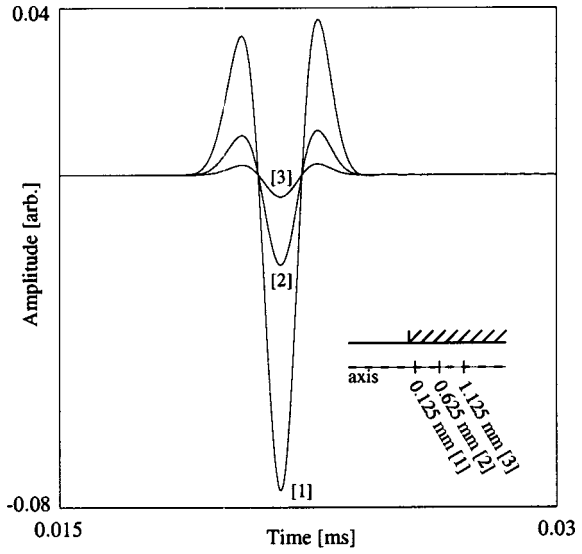


FIG. 14. The time trace showing the displacements in the rigid section of the cylinder at three locations: 0.125, 0.625, and 1.125 mm away from the interface.

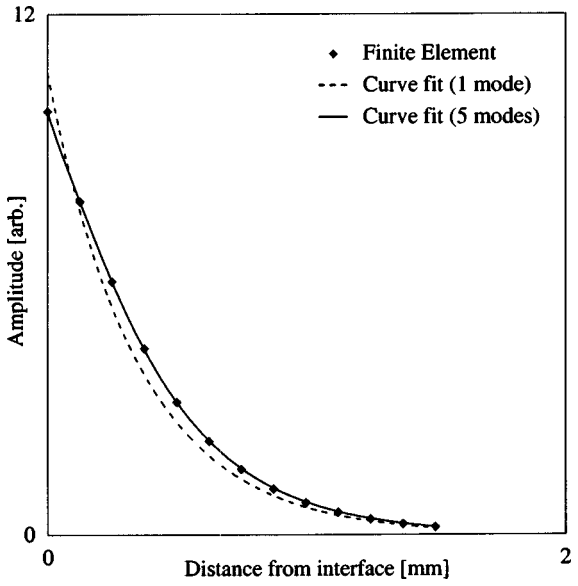


FIG. 15. Modulus of the displacement at 0.2 MHz. Details regarding the curve fits are given in the text.

nonpropagating L(0,1) in the rigid section were included in the modal solution.

Zemanek[16] has already pointed out that inhomogeneous modes are always created in pairs with an equal imaginary wave number but with real wave numbers of opposite sign. This way, a spatially attenuated standing wave is set up that does not propagate energy. As an example for the free cylinder, L(0,2) and L(0,3) represent such a pair of inhomogeneous modes. Indeed, it was found in the calculations that the scattering parameters, *b*, of the corresponding inhomogeneous modes are identical to a high degree of accuracy.

The displacements in the *z* direction on the axis of the cylinder with the rigid surface were also monitored in the FE analysis. They can be thought of as the superposition of the displacements of all the nonpropagating and inhomogeneous modes excited at the interface. Figure 14 shows the displacements as a function of time at three different distances from the interface, which is located at $z=0$ mm. It can clearly be seen that there is no propagation, and the amplitude decreases with the distance from the interface. These modes can therefore be thought of as local vibrations with a spatial attenuation away from the discontinuity. The decay of the displacement amplitude at 0.2 MHz, obtained from the Fourier transform of the signal, with distance from the interface is plotted in Fig. 15. As the total displacement, *u*, in the rigid part of the waveguide is a superposition of all the spatially decaying nonpropagating and inhomogeneous modes involved, the *z* dependence of the modulus can be expressed in the form

$$u(z)=\sum_{m'} u_{m'}(z)=\sum_{m'} |u_{m'}|e^{-i\xi_{m'}z}e^{-\alpha_{m'}z}, \tag{72}$$

where ξ is the real wave number, and α is the attenuation. Here, the time dependence has been omitted. Since the inhomogeneous modes occur in pairs with real wave numbers of opposite sign, this expression becomes

$$u(z)=\sum_{m'} |u_{m'}|\cos(\xi_{m'}z)e^{-\alpha_{m'}z}. \tag{73}$$

For nonpropagating modes, the real wave number is zero, and the cosine takes the value unity. The values for the real wave number, ξ, and the attenuation, α, are obtained from DISPERSE, and a curve can be fitted to the FE results using the above equation, with the $|u_{m'}|$'s being the fitting parameters. It is found that, even though the modal solution gives accurate results when including only one non-propagating mode, the decay function cannot be satisfactorily fitted using only one mode (see Fig. 15). Instead, a minimum of five nonpropagating and inhomogeneous modes is needed to obtain a precise fit. This means that, although the modal solution gives an accurate representation of the reflection coefficient when including a minimal number of modes, more modes might actually be significantly involved in the scattering process.

B. Scattering from epoxy loading

The different examples presented below are related to the application of a guided wave system to the cure monitoring of epoxy resins. A steel cylinder that is partly embedded in an epoxy resin is considered. The waveguide and the epoxy resin are modeled as purely elastic materials.

Figure 16 shows the modulus of the amplitude reflection coefficient for the lowest-order longitudinal mode, L(0,1), in a 1 mm radius steel wire using the material properties given in Table I and using just the lowest-order longitudinal mode in the embedded section in the modal solution. In the same plot, the FE results are shown, and the agreement is very good. Including more inhomogeneous modes in the free section and the equivalent modes in the embedded section changes the reflection coefficient only by a negligible amount. Therefore, the influence of these modes on the reflection coefficient is minimal in this case. This is in accordance with the findings of Engan in the case of the scattering

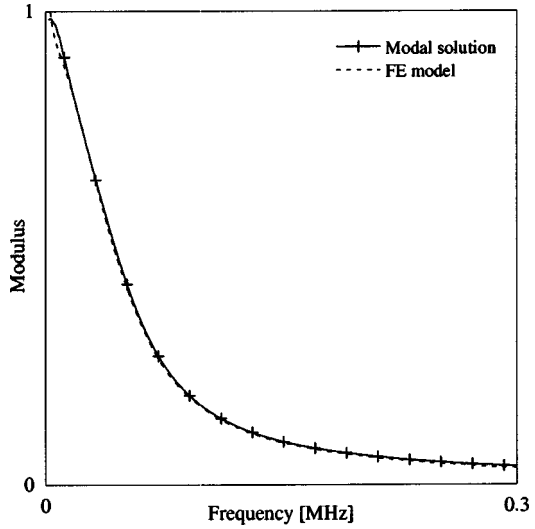


FIG. 16. The modulus of the amplitude reflection coefficient of L(0,1) in a 1 mm radius steel bar embedded in epoxy. The material parameters for this analysis are given in Table I. The plot shows the results obtained with the modal solution and the Finite Element (FE) model.

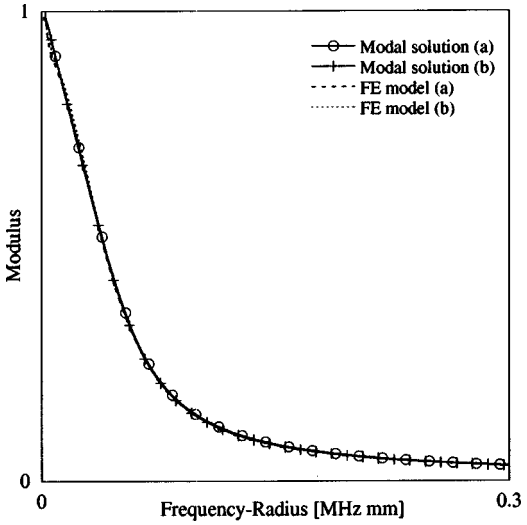


FIG. 18. The dependence of the modulus of the reflection coefficient on the longitudinal velocity for two different materials, "a" and "b," respectively. The material properties used in the analysis are given in Table I.

of torsional modes from a step change in diameter,[19] which suggest that only a few non-propagating modes are needed for a reasonably accurate description. The error in the power flow calculation was found to be of the order of $\epsilon \approx 10^{-4}-10^{-5}$, even with small numbers of inhomogeneous modes.

It is interesting to note that the amplitude reflection coefficient curve scales with the product of the frequency and the radius of the cylinder (see Fig. 17). This is not surprising as the reflection coefficient is calculated using the mode shapes, which, like the dispersion curves, scale with the frequency–radius product. This has been confirmed both with FE modeling and using the modal solution, although only the FE results are plotted in Fig. 17.

The power of the L(0,1) mode at low-frequency–radius products is distributed fairly evenly over the cross section of the waveguide. However, the scattering occurs on the circumference of the waveguide. It seems reasonable that under these conditions the ratio of circumference to the cross-sectional area roughly relates to the ratio of the scattered to the total power of the incident mode. It was found that in this case only the lowest-order longitudinal mode plays a significant role in the scattering process, and therefore the ratio of scattered to total power governs the reflection coefficient. Because of the geometry, this ratio decreases with increasing radius. This explains why the reflection coefficient is large for low-frequency–radius products, and falls off to low values at higher frequency-radius products.

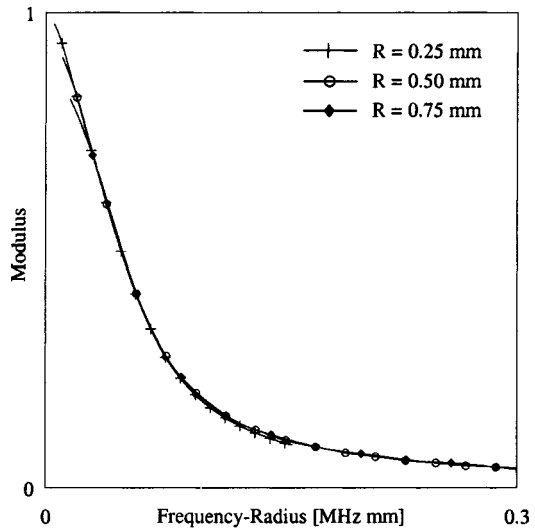


FIG. 17. The modulus of the amplitude reflection coefficient of L(0,1) in a steel bar embedded in epoxy as a function of the product of frequency and radius. The material parameters for the analysis are given in Table I. The plot shows the Finite Element results obtained with three different wire radii *R*.

The particle displacements of L(0,1) at very low frequencies at the surface of the waveguide are mainly in the axial direction. Therefore, guided wave attenuation is mostly caused by the leakage of shear waves and depends strongly on the shear velocity of the embedding material, and only to a small extent on its longitudinal velocity. This is illustrated in Fig. 18, where the reflection coefficient for embedding materials with constant shear velocity and density but a 100% difference in the longitudinal velocity are shown. There is a negligible difference in the curves for the two materials. For a cure monitoring application, the change in shear properties with curing time is significant, and a significant change in the reflection coefficient is expected. Figure 19 shows the amplitude reflection coefficient for different values of shear velocity, calculated using both the modal solution and FE modeling. The reflection coefficient increases with increasing shear velocity, since the difference in surface impedance between the free and the embedded cylinder is larger. Therefore the reflection coefficient can be used to evaluate the shear properties of an embedding material.

VII. CONCLUSIONS

The scattering of guided waves in a free bar from the point where the waveguide enters an embedding material has

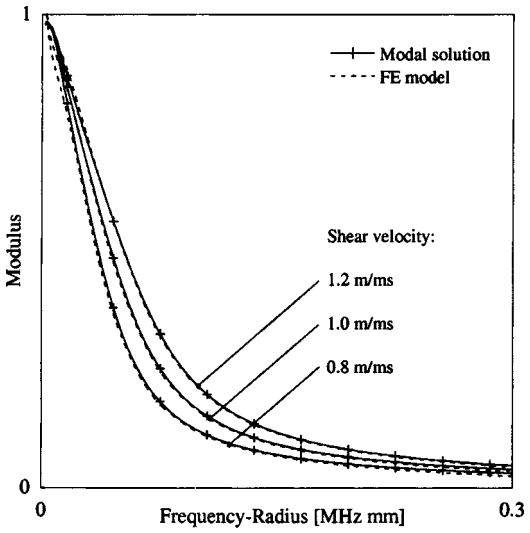


FIG. 19. The modulus of the amplitude reflection coefficient of L(0,1) in a steel bar for different values of epoxy shear velocity. All other material properties of the epoxy are according to Table I.

been discussed. It has been shown that for a wave that is incident from the free section of the bar, a scattering formula equivalent to one derived using the S parameter formalism is obtained by matching the fields at the interface between the free and the embedded waveguide. A modal solution shows good agreement with results obtained from Finite Element modeling.

For the modal solution presented here, orthogonality of the leaky modes in the embedded waveguide is assumed and appears to be satisfied sensibly in the example that is studied. However, further research should investigate an orthogonality relationship for leaky waveguide modes. The mathematical form of such a relationship, for example whether an integral form includes the embedding half-spaces or only the waveguide cross section, determines the modal solution in the case where a guided wave is incident from the embedded section of the waveguide.

Orthogonality is given in the case where the embedding material is perfectly rigid. The results for this case show that, although several nonpropagating and inhomogeneous modes are involved in the scattering process, their influence on the modal solution is minimal.

The nature of the reflection coefficient when the embedding material is an epoxy resin has been discussed in detail. It has been shown that the reflection coefficient is a function of the frequency–radius product, as are the dispersion relationships. It was found that, in the frequency range that was covered by the analysis, the reflection coefficient depends only minimally on the longitudinal velocity of the resin, but is a strong function of the shear velocity. As in the case where the embedding material is rigid, the influence of the nonpropagating and inhomogeneous modes was found to be minimal.

ACKNOWLEDGMENTS

The authors gratefully acknowledge helpful discussions with Professor Richard Weaver, University of Illinois.

[1] T. Vogt, M. Lowe, and P. Cawley, "Cure monitoring using ultrasonic guided waves in wires," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, Melville NY, 2001), Vol. 20B, pp. 1642–1649.
[2] E. P. Papadakis, "Monitoring the moduli of polymers with ultrasound," J. Appl. Phys. **45**, 1218–1222 (1974).
[3] B. N. Pavlakovic, M. J. S. Lowe, D. N. Alleyne, and P. Cawley, "DISPERSE: A general purpose program for creating dispersion curves," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1997), Vol. 16, pp. 185–192.
[4] B. N. Pavlakovic and M. J. S. Lowe, "A general purpose approach to calculating the longitudinal and flexural modes of multi-layered, embedded, transversely isotropic cylinders," in Ref. 3, Vol. 18, pp. 239–246.
[5] A. H. Nayfeh and P. B. Nagy, "General study of axisymmetric waves in layered anisotropic fibers and their composites," J. Acoust. Soc. Am. **99**, 931–941 (1996).
[6] T. Vogt, M. Lowe, and P. Cawley, "Ultrasonic waveguide techniques for the measurement of material properties," in Ref. 1, Vol. 21, pp. 1742–1749.
[7] J. O. Kim and H. H. Bau, "Instrument for simultaneous measurement of density and viscosity," Rev. Sci. Instrum. **60**, 1111–1115 (1989).
[8] R. D. Costley, K. Balasubramaniam, W. M. Ingham, J. A. Simpson, and V. Shal, "Torsional waveguide sensor for molten materials," in Ref. 3, Vol. 17, pp. 859–866.
[9] D. N. Alleyne, B. Pavlakovic, M. J. S. Lowe, and P. Cawley, "Rapid, long range inspection of chemical plant pipework using guided waves," Insight **43**, 93–96 (2001).
[10] M. D. Beard, M. J. S. Lowe, and P. Cawley, "Development of a guided wave inspection technique for rock bolts," Insight **44**, 19–24 (2002).
[11] J. O. Kim, H. H. Bau, Y. Liu, L. C. Lynnworth, S. A. Lynnworth, K. A. Hall, S. A. Jacobson, J. A. Korba, R. J. Murphy, M. A. Strauch, and K. G. King, "Torsional sensor applications in two-phase fluids," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 563–574 (1993).
[12] B. A. Auld, *Acoustic Fields and Waves in Solids* (Krieger, Malabar, FL, 1990), Vol. 2.
[13] P. J. Torvik, "Reflection of wave trains in semi-infinite plates," J. Acoust. Soc. Am. **41**, 346–353 (1967).
[14] J. J. Ditri, "Some results on the scattering of guided elastic SH waves from material and geometric waveguide discontinuities," J. Acoust. Soc. Am. **100**, 3078–3087 (1996).
[15] E. Le Clézio, M. Castaings, and B. Hosten, "Réflexion d'ondes de Lamb à l'extrémité de plaque: solution analytique et mesures," in *Proceedings of the 5th Congrès Français d'Acoustique*, 2000, pp. 177–180.
[16] J. Zemanek, "An experimental and theoretical investigation of elastic wave propagation in a cylinder," J. Acoust. Soc. Am. **51**, 265–283 (1972).
[17] J. J. Ditri, "Utilization of guided elastic waves for the characterization of circumferential cracks in hollow cylinders," J. Acoust. Soc. Am. **96**, 3769–3775 (1994).
[18] H. Bai, A. H. Shah, N. Popplewell, and S. K. Datta, "Scattering of guided waves by circumferential cracks in steel pipes," J. Appl. Mech. **68**, 619–631 (2001).
[19] H. E. Engan, "Torsional wave scattering from a diameter step in a rod," J. Acoust. Soc. Am. **104**, 2015–2024 (1998).
[20] D. N. Alleyne, M. J. S. Lowe, and P. Cawley, "The reflection of guided waves from circumferential notches in pipes," J. Appl. Mech. **65**, 635–641 (1998).
[21] M. J. S. Lowe, D. N. Alleyne, and P. Cawley, "The mode conversion of a guided wave by a part-circumferential notch in a pipe," J. Appl. Mech. **65**, 649–656 (1998).
[22] D. N. Alleyne and P. Cawley, "The interaction of Lamb waves with defects," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 381–397 (1992).
[23] D. C. Gazis, "Three dimensional investigation of the propagation of waves in hollow circular cylinders," J. Acoust. Soc. Am. **31**, 568–578 (1959).
[24] M. G. Silk and K. F. Bainton, "The propagation in metal tubing of ultrasonic wave modes equivalent to Lamb waves," Ultrasonics **17**, 11–19 (1979).
[25] P. Kirrmann, "On the completeness of Lamb modes," J. Elast. **37**, 39–69 (1994).
[26] M. J. S. Lowe, "Matrix techniques for modeling ultrasonic waves in multilayered media," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 525–542 (1995).
[27] B. N. Pavlakovic and M. J. S. Lowe, *Disperse Manual—A System For*

Generating Dispersion Curves, NDT Laboratory, Imperial College; website www.me.ic.ac.uk/dynamics/ndt, 2001.
[28] J. E. Murphy, G. Li, and S. A. Chin-Bing, ''Orthogonality relation for Rayleigh–Lamb modes of vibration of an arbitrarily layered elastic plate with and without fluid loading,'' J. Acoust. Soc. Am. **96**, 2313–2317 (1994).
[29] R. Briers, O. Leroy, G. N. Shkerdin, and Y. V. Gulyaev, ''Mode theory as a framework for the investigation of the generation of a Stonely wave at a liquid–solid interface,'' J. Acoust. Soc. Am. **95**, 1953–1966 (1994).
[30] M. Hayes and M. J. P. Musgrave, ''On energy flux and group velocity,'' Wave Motion **1**, 75–82 (1979).
[31] M. Onoe, H. D. McNiven, and R. D. Mindlin, ''Dispersion of axially symmetric waves in elastic rods,'' J. Appl. Mech. **29**, 729–734 (1962).
[32] D. Hitchings, *FE77 User Manual* (Imperial College, London, 1994).

Continuation of acoustic near-fields

Earl G. Williams[a)]
Code 7130, Naval Research Laboratory, Washington, DC 20375

(Received 15 April 2002; revised 9 October 2002; accepted 28 October 2002)

This paper deals with the analytic continuation of a coherent pressure field specified on a finite sheet located close to and conformal to the surface of a vibrator. This analytic continuation is an extension or extrapolation of the given (measured) field into a region outside and tangential to the original finite sheet, and is based on the Green's function (the transfer function) relating acoustic quantities on the two conformal surfaces. The continuation of the measured pressure field is an inverse problem that requires the use or regularization theory, especially when noise is present in the data. An iteration algorithm is presented that is successful in continuing the pressure field into the tangential sheet. The results are accurate close to the original boundary and taper (decay) toward zero with distance away from it. The algorithm is tested on numerical and experimental data from a point-driven rectangular plate. Results show the successful extrapolation (continuation) of this data into an area nearly double that of the original pressure field. This algorithm is not limited to planar surfaces and can be applied to arbitrarily shaped surfaces. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1528173]

PACS numbers: 43.20.Tb, 43.20.Ye, 43.40.Rj [MO]

I. INTRODUCTION

Measurements of acoustic fields on a sheet-like surface are by their very nature limited in physical extent. For example, medical imaging uses arrays that intercept scattered fields over a finite (and discretized) surface. In acoustics, farfield scattering measurements are made over a limited solid angle from which information about the scatterer is extracted either directly or using inverse methods. Near-field acoustical holography (NAH) uses microphone array measurements in the extreme near-field of the source to solve an inverse problem—reconstruction of the normal velocity of the vibrator located nearby. In almost all cases, if the measurement apertures are increased in size, more accurate information is gathered about the source of interest. However, larger measurement apertures may not be economical and an interesting question arises as to whether or not the given pressure can be continued mathematically on the tangent to the sheet, effectively increasing the size of the measurement aperture without further measurement. In particular, we are interested in the case, typical of NAH, in which an array of pressure measurements are made close to, and equidistant from (conformal to), a vibrating surface. In applications of NAH the pressure data is zero padded (a gross form of extrapolation) and a spatial Tukey window[1] is used to taper the measured data at the boundary into the null extension of the field. This padding, however, provides no new information and compromises the measured data near the boundary due to the taper. If instead the zero padded region can be filled (extrapolated), with data that reflects the physics of the measured pressure and transitions smoothly to the measured field, then the Tukey window is not needed and a compromise of the measured pressure field can be avoided.

The continuation of an analytic field has a rich history in electrostatics—in solutions of the Laplace equation, centering around analytic continuation. In the realm of the Helmholtz wave equation, analytic continuation often involves series expansions in terms of radiating harmonics of the field using spherical harmonics and spherical Bessel functions, and a body of literature surrounding the Rayleigh hypothesis[2,3] deals with the question of the validity of these expansions. It is pointed out that particular series expansions fall when they approach the singularities of the given scatterer (the vibrating surface). These singularities can be identified for simple cases.[4,5] It is not within the scope of the present paper to determine the range of validity of the expansions used here, and we assume that we do not cross any singularities in the continuation of the measured field, an assumption that is trivial for the examples provided in Sec. III.

It is not evident at first, but this mathematical extension of the measured data is an inverse problem requiring regularization due to noise in the measurement. Furthermore, there is one and only one exact extrapolation of the given pressure field. In the exact sense the extrapolation is unique. The iteration procedure presented here attempts to reproduce this unique field outside, with an accuracy that one would expect, given the nature of inverse problems, that degrades with distance away from the edge of the given data. We will show in examples that the degradation is ''graceful,'' tapering to zero. This aperture extension has proven to be very useful in developing a general patch NAH algorithm that uses the limited (on a patch) pressure array information to reconstruct the velocity field on the vibrator directly below the patch. However, patch NAH itself will be the subject of a future paper. One issue, however, is fundamental. Dealing with evanescent waves is an integral part of the aperture extrapolation procedure, since they provide enhanced resolution for the velocity reconstruction, and are critical to the near-field radiation problem.

Iteration procedures have appeared in the leterature deal-

a)Electronic mail: williams@pa.nrl.navy.mil

ing with NAH. For example Landweber iteration was used by Ih[6] and Williams[7], and recursive Weiner filtering by Bai[8]. These procedures, however, do not extend or continue the measurement aperture. The iteration procedure that we present here is similar to the one presented to calculate the radiation from unbaffled planar vibrators.[9] In that paper a given velocity field, initially baffled, is used to calculate the surface pressure field (via Rayleigh's integral) that is zeroed out (due to the unbaffled assumption) outside the vibrator and the resulting field used to recompute the normal velocity (via an inversion of Rayleigh's integral). The center is replaced with the original velocity and the procedure is repeated. A similar approach has been published recently,[10] which deals with planar vibrators and approaches the problem in wave number space. In this case the wave number spectrum of the measured pressure field (zero-padded initially) is zeroed out beyond a predefined *k*-space cutoff and used to recompute the pressure. The pressure within the measured area is replaced with the original pressure and the procedure is repeated. We build upon this latter approach and also extend it to arbitrarily shaped vibrators. In particular, we avoid the need to specify, *a priori*, the *k*-space cutoff by using regularization theory based on the physics of the radiation process. Furthermore, we add a taper to the *k*-space filter resulting from the regularization that increases the accuracy of the procedure.

II. THEORY

Let x be the two-dimensional coordinate of a point on a finite sheet Ω_i representing the measurement aperture and let the *tangential* extension of this finite sheet be denoted by Ω_0. Define the full aperture $\Omega := \Omega_i \cup \Omega_0$. Let $p(x)$ for $x \in \Omega$ be the exact pressure field on the full aperture. Our intention is to develop an iteration algorithm that attempts to determine $p(x)$ for $x \in \Omega_0$ given only $p(x)$ for $x \in \Omega_i$ and a model for the radiation process. This model includes the known location of the conformal vibrating source located close to the measurement surface. The measured pressure [which contains spatially uncorrelated random noise $\epsilon(x)$ with variance σ] is defined by $p^{\delta}(x) := p(x) + \epsilon(x)$. All quantities with noise will be indicated by the δ superscript. In order to exploit the uniqueness of the extrapolation we require a system model relating geometrical information about the two parallel (conformal) surfaces. This model is provided by the transfer function between the (unspecified) normal velocity $\dot{w}(y)$ on the vibrating surface and the measured pressure $p(x)$ and its extension. Here y is the coordinate on the conformal vibrating surface defined by $\dot{\Omega}$ (which is located by a normal projection of the points $x \in \Omega$ to $y \in \dot{\Omega}$). The system model is

$$p(x_i) = \mathbf{H}\dot{w}(y_j), \quad \{x_i \in \Omega, y_j \in \dot{\Omega}\}, \tag{1}$$

where $1 \leqslant i, j \leqslant M$ and M is the total number of points on each surface (assumed equal). Thus $p(x_i)$ and $\dot{w}(y_j)$ are column vectors resulting from the discretization of the field spanning each of the two surfaces, respectively, and the transfer function $\mathbf{H}$ is an $M \times M$ matrix. For example, the system model for planar surfaces would be provided by a discretization of Rayleigh's first integral,[1] leading to the transfer function $\mathbf{H}$. For more general surfaces, the discretization and transfer function are provided by a boundary element (BEM) approach.[11–13] We emphasize that even though the results in this paper deal with planar boundaries, the starting point of Eq. (1) is quite general and can be obtained from an application of the Helmholtz integral equation (HIE) to arbitrary shaped surfaces through this BEM discretization. In these approaches a second HIE is written for the pressure field on the surface so that the surface pressure can be eliminated, thus leading to Eq. (1). As long as the measurement patch is conformal to this surface, the conclusions of this paper are directly applicable to the arbitrary surface case.

The iteration procedure we develop starts with the measured pressure zero padded to fill out the extended aperture:

$$\hat{p}(x_i) = \begin{cases} p(x_i) & x_i \in \Omega_i \\ 0 & x_i \in \Omega_0 \end{cases}. \tag{2}$$

Successive iterations attempt to smooth the pressure field from the bounding edge of Ω_i into the exterior region Ω_0. This smoothing is done using a low pass wave number filter F^{α} (the parameter α that appears in the superscript represents the break point of the filter) operating on a wave number decomposition of $\hat{p}(x_i)$, $x_i \in \Omega$. This wave number decomposition is given by $\mathbf{U}^H \hat{p}(x_i)$, where $\mathbf{U}^H$ is a unitary matrix with each row an orthonormal basis function. $P(k_i) := \mathbf{U}^H \hat{p}(x_i)$ provides a column vector $P(k)$ of the amplitudes of the *k*-space components (wave number or Fourier decomposition) of $\hat{p}(x)$. We will see that $\mathbf{U}^H$ has the important property that as one moves down its rows, the number of zero crossings of the mode (orthonormal basis function contained in the row) generally increase. This corresponds to $|k|$ increasing so that the radiation becomes more and more evanescent as we move down the rows. By evanescent we mean that the waves tend to decay exponentially in the direction normal to the vibrating surface $\dot{\Omega}$. Thus the rows of $\mathbf{U}^H$ are ordered from least evanescent (smallest wave number at the top row) to most evanescent waves (largest wave number at the bottom row).

When data with noise is considered, then the amplitudes $P^{\delta}(k) := \mathbf{U}^H \hat{p}^{\delta}(x)$ on Ω for the largest values of $|k|$ correspond *only to noise*, under the assumption that the evanescent waves generated on the vibrating surface decay so rapidly for these values that the evanescent wave amplitudes have dropped below the noise level in the pressure data. These evanescent waves cannot be observed on Ω. In an attempt to filter out this noise and at the same time to pass the evanescent (and nonevanescent) waves on Ω that are above the noise, we design a low-pass *k*-space filter F^{α}. The shape as well as the break point of this filter (defined when the filter amplitude is 0.5) is critical. Its taper depends on the system model (which encodes the physics of the radiation mechanism for evanescent waves between $\dot{\Omega}$ and Ω) and as $|k|$ increases should match and slightly exceed the increase in the decay of the evanescent waves.

There is another set of high-wave number components that are deleterious to the process, and that are not part of the physics of the vibrator. They are created by the sharp transi-

tion in $\hat{p}$ from Ω_i to Ω_0 due to the initial zero padding. Thus the low-pass filter also filters out much of the spurious high-wave number components created initially by zero padding.

The filter is given on the diagonal of the diagonal matrix F^α. Thus, the filtered wave number spectrum of the zero-padded pressure is given by $F^\alpha \mathbf{U}^H \hat{p}^\delta$. We will discuss later how the filter and its break point are determined. Finally the filtered (and extrapolated) pressure in real space is given by

$$p^{\alpha,\delta} = \mathbf{U} F^\alpha \mathbf{U}^H \hat{p}^\delta, \tag{3}$$

where the extra superscript α is added to the pressure to indicate its dependence on the filter shape and cutoff. Whereas $\hat{p}^\delta$ is null in Ω_0, $p^{\alpha,\delta}$ will be nonzero due to the smoothing of the pressure jump on the boundary. This smoothing of the zero-padded data provides the first estimate of the extrapolation of the measured field.

Two important issues must be quantified before we can develop an iteration procedure using Eq. (3). The first is the shape of the filter and the second is the cutoff. To derive the shape we need a model of the radiation process since the taper, as we mentioned above, should match the decay with an increasing wave number of the evanescent waves, as seen on Ω, which arise from the actual vibration of the source. The amount of decays is encoded in the Green's function (that is, the transfer function) relating the velocity on the vibrating surface to the pressure on the measurement surface. We can extract this decay for general surfaces, at the same time determining the basis function set for the pressure using a singular value decomposition[14–17] of the transfer function:

$$\mathbf{H} = \mathbf{U} \mathbf{G} \mathbf{V}^H, \tag{4}$$

where $\mathbf{G}$ is a diagonal matrix of singular values providing the decay rates of the evanescent waves. It is ordered so that the first element is the least evanescent (often corresponding to the DC wave number) and the last element is the most evanescent (smallest singular value). Furthermore, rows of $\mathbf{U}^H$ contain the modal decomposition (orthonormal basis),[18] that we need to translate the pressure $\hat{p}^\delta(x_i)$, $x_i \in \Omega$ into k space, as discussed above.

For planar geometries, which we will draw upon later for the examples, there are two approaches to generating Eq. (4). Besides the SVD approach one can recast $\mathbf{U}^H$ and $\mathbf{V}^H$ as discrete Fourier transforms, as discussed recently.[7] This allows the use of Fourier acoustics[1] and an accompanying wealth of tools, not the least of which is the fact that the discretized *analytic* version of $\mathbf{G}$ is known in this case. Also k space is given simply by two-dimensional Fourier transforms approximated by the discrete Fourier transform (DFT). We will discuss this in more detail in Sec. III. In the analysis below, both of these two approaches, SVD or DFT, use the exact same formulation and for planar or cylindrical geometries either approach can be taken. We find that the DFT approach is significantly faster, however, although it cannot be used for general geometries.

A. Construction of the wave number filter

Returning to the filter F^α, we construct the shape of the filter for some given break point α. Tikhonov regularization theory provides one possible estimate of the shape of this filter. We have found that this theory is very robust and incorporates the physics of the radiation process. Since there is a large body of literature[19] dealing with Tikhonov regularization and a recent paper[7] pertains directly to the above analysis, we will not repeat the details here. Instead we extract the derived filter [see Eq. (56) of Williams[7]]:

$$F^\alpha := \mathrm{diag}\left(\cdots \frac{|\lambda_i|^2}{|\lambda_i|^2 + \alpha(\alpha/\alpha + |\lambda_i|^2)^2} \cdots \right), \quad 1 \leqslant i \leqslant M, \tag{5}$$

where λ_i are the ordered singular values given in Eq. (4) by

$$\mathbf{G} = \mathrm{diag}(\lambda_1, \ldots, \lambda_M), \quad \lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_M, \tag{6}$$

and M is the number of points in Ω. Thus, we can see that the singular values (decay rates of the evanescent waves) determine the shape of the filter for a given value of α. Note that when $|\lambda_1|^2 \gg \alpha$ the filter unity (full pass) and when $|\lambda_M|^2 \ll \alpha$ it is zero (all components blocked) so it is truly a low-pass filter.

Next, we need to determine the break point ($F^\alpha = 0.5$) of the filter that will depend upon α. Clearly this must be dependent of the random noise in the measurement and we need a model that incorporates the noise level. This model is provided by a simple hypothesis: assuming that the filter is correctly tapered so that it matches the decay with an increasing wave number of the evanescent waves from $\dot{\Omega}$ and thus is completely successful in removing the random noise in the measurement, then the following statement must be true for $x_i \in \Omega$:

$$\| p^{\alpha,\delta} - p^\delta \| \sqrt{M} \approx \| p - p^\delta \| / \sqrt{M} = \sigma, \tag{7}$$

where the latter equality is just the definition of the variance of the noise σ since p is the exact pressure, and $\| \cdot \|$ is the L_2 norm. The approximation symbol above indicates the relative success of the filtering of the random noise. This intuitive equation is attributed to Morozov and is called the Morozov discrepancy principle.[20] Assuming equality and using Eq. (3), the first term of Eq. (7) becomes

$$\| (\mathbf{U} F^\alpha \mathbf{U}^H - I) p^\delta \| / \sqrt{M} = \sigma. \tag{8}$$

Given Eq. (5) and a knowledge of σ, one can solve this equation for α. It turns out that Eq. (8) is monotonic in α,[21,7] so simple procedures can be used to solve for α.

However, the implementation of Eq. (8) requires an estimate of the noise variance σ in the data. We obtain this using a procedure similar to Eq. (8). Given that the taper of the filter matches the evanescent wave decay with increasing wave number, then the highest wave numbers above the break point correspond only to noise in the data. Thus, the expected value $\mathcal{E}$ of the last few basis vectors operating on the noise $\boldsymbol{\epsilon}$ yields the following equation:[19]

$$\mathcal{E}(|U_q^H p^\delta|) \approx \mathcal{E}(|U_q^H \boldsymbol{\epsilon}|) = \sigma, \quad \text{for small } |\lambda_q|, \tag{9}$$

where U_q^H is the qth row of $\mathbf{U}^H$. We average over the last Q basis vectors,

$$\| U_q^H p^\delta \| / \sqrt{Q} \approx \sigma, \quad M - Q + 1 \leqslant q \leqslant M, \tag{10}$$

for an approximate implementation of the expectation, assuming that the evanescent waves associated with the last Q

modes have all dropped below the noise. For more details of the above procedures and equations see Ref. 7.

III. ITERATION PROCEDURE

Returning to the first estimate of the extrapolated pressure field, provided by Eq. (3), we note that due to the smoothing at the interface $p^{\alpha,\delta}$ will not match the measured field p^{δ} in $x \in \Omega_i$ particularly near the interface. Thus the iteration procedure is built around the subsequent replacement in Ω_i of $p^{\alpha,\delta}$ with the measured field, a technique used previously by Williams[9] to deal with the radiation from unbaffled vibrators and Saijyou[10] for data extrapolation. Thus, the extrapolated pressure field for the first iteration is set to

$$\hat{p}^{(1,\alpha,\delta)} := \begin{cases} p^{\delta} & x \in \Omega_i \\ p^{(1,\alpha,\delta)} & x \in \Omega_0 \end{cases}, \tag{11}$$

where the caret over the quantity indicates that the smoothed field has had the center replaced with the original data. The smoothed pressure $p^{(1,\alpha,\delta)}$ resulted from the filtering process of Eq. (3) applied to the zero padded, initial value $\hat{p}^{(0,\delta)}$ on the right-hand side. This zero-padded initial value (not dependent on α) was given in Eq. (2) and is now

$$\hat{p}^{(0,\alpha,\delta)} = \hat{p}^{(0,\delta)} := \begin{cases} p^{\delta} & x \in \Omega_i \\ 0 & x \in \Omega_0 \end{cases}. \tag{12}$$

Thus, the iteration steps, recasting Eq. (10), Eq. (8), Eq. (3), and Eq. (11), respectively, are

$$\text{(I1)} \quad \|\mathbf{U}_q^H \hat{p}^{(i,\alpha^i,\delta)}\| / \sqrt{Q} \approx \sigma^i \Rightarrow \sigma^i,$$

$$\text{(I2)} \quad \|(\mathbf{U} F^{\alpha^{i+1}} \mathbf{U}^H - I)\hat{p}^{(i,\alpha^i,\delta)}\| / \sqrt{M} = \sigma^i \Rightarrow \alpha^{i+1},$$

$$\text{(I3)} \quad p^{(i+1,\alpha^{i+1},\delta)} = \mathbf{U} F^{\alpha^{i+1}} \mathbf{U}^H \hat{p}^{(i,\alpha^i,\delta)} \Rightarrow p^{(i+1,\alpha^{i+1},\delta)},$$

$$\text{(I4)} \quad \hat{p}^{(i+1,\alpha^{i+1},\delta)} = \begin{cases} p^{\delta} & x \in \Omega_i \\ p^{(i+1,\alpha^{i+1},\delta)} & x \in \Omega_0 \end{cases} \Rightarrow \hat{p}^{(i+1,\alpha^{i+1},\delta)},$$

where the right-hand side indicates the quantity that is calculated in each step.

We start with $i=0$ and $\hat{p}^{(0,\delta)}$ from Eq. (12) and solve for the noise estimate σ^0 using iteration equation (I1). Then σ^0 is used in (I2) with F^{α^1} given by Eq. (5) to solve for α^1 (using a binary subdivision method on α). The shape and break point of the filter are now completely defined. Using this filter, a smoothed version of the pressure field is obtained from (I3). This smoothed version provides the first estimate of the extrapolation of p^{δ}, $x \in \Omega_i$ into Ω_0. Finally, (I4) replaces the center with the initial measured pressure and we continue with (11) once again (setting $i=2$).

Note the SVD decomposition needed to define $\mathbf{U}$ and to construct F^{α} is only done once, thus $\mathbf{U}$ and λ_i do not change from iteration to iteration. The iteration procedure is stopped when there is minimal change in the updated smoothed pressure, that is,

$$\|p^{(i+1,\alpha^{i+1},\delta)} - p^{(i,\alpha^i,\delta)}\| / \sqrt{M} < 0.1\sigma^i. \tag{13}$$

This equation reflects the fact that the minimal pressure change must depend on the noise level in the smoothed data. The *ad-hoc* factor of 0.1 worked well in most cases. However, we found it helpful to require a certain minimum number of iterations to take place before any cutoff criterion is applied. Unfortunately we cannot present proof that the iteration procedure presented above converges, but have found in all the test cases explored that it is robust and always converges as long as we use Eq. (13) and an *ad-hoc* factor that is not too small. If the *ad-hoc* factor is too large, the iteration procedure cuts off too soon and the extrapolation is less accurate. We now show some examples using synthetic data. We will discuss the question of uniqueness of the extrapolated field in Sec. V.

IV. NUMERICAL EXAMPLE: POINT-DRIVEN CLAMPED PLATE

We use an analytic model for a center-driven clamped plate, driven with a constant force over frequency. The pressure field is calculated in the extreme near-field 1.0 cm from the plate surface using an adaptive quadrature algorithm applied to the Rayleigh's first integral. The point-driven baffled steel plate was rectangular, 0.4×0.34 m, with a thickness of 2 mm. The pressure field was computed on a gird 81×61 points with a lattice spacing in both directions of 1.0 cm. Thus, the hologram covers an area of 0.81×0.61 m. For the studies below we added random noise to this hologram pressure so that the signal-to-noise ratio defined by $\|p\|_2/(\sigma\sqrt{M})$ was 40 dB. From this hologram a 11×11 patch of points is selected over the vibrating area at a specified frequency. Our objective is to apply the iteration algorithm to this patch to continue the pressure field to the region outside the patch (in the same plane) and compare this extrapolation to the known field there.

We choose examples that are typical of the results of many studies, some that included real hologram data. Results with experimental data are presented in the next section. The first two examples (each at 2100 Hz) using the synthetic data in Figs. 1 and 2 are for two 11×11 patches located in different areas in the hologram plane, but both located within (projecting normally) the boundary of the plate. The patches are zero padded to fill out an aperture of 51×51 pixels. All the processing in iteration equations (11)–(14) are carried out using the 51×51 point aperture and DFT processing.

Figures 1 and 2 show the real part of $p^{(i,\alpha^i,\delta)}(x)$ for $i=1$, 500 and the final value with gray scale encoding of the amplitude. The imaginary part is similar and is not shown. The first panel in each figure shows $p^{\delta}(x)$ with a SNR of 40 dB, the known field for a comparison. The progressive extrapolation of the patch data can be seen in the bottom two panels. Upon comparison to the top left panel, it is apparent that the extrapolated pressure field is only accurate very close to the starting patch. As one moves away from the boundary the extrapolation fades to a zero pressure field. The normalized L_2 errors E, defined by

$$E = 100\|p_{\text{exact}} - p_{\text{extrapolated}}\|_2 / \|p_{\text{exact}}\|_2, \tag{14}$$

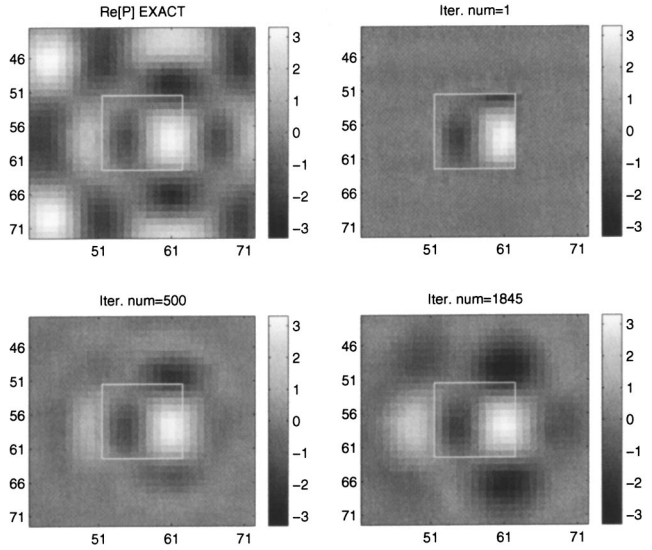


FIG. 1. An 11×11 patch (square white box indicates its boundaries) is extracted from the exact hologram, shown at top left, at 2100 Hz. The pressure field is shown at the beginning of the iteration procedure (Iter. num=1) in the top right panel, and after 500 iterations on the bottom left, and finally at the end after 1845 iterations. The real part of the pressure field is displayed with a gray scale as shown (pascals per Newton force). Only the center 32×32 region out of the 51×51 full aperture is shown. DFT processing is used.

for the pressure in five concentric square rings one point wide, moving away from the patch, were 4.8%, 14.7%, 30.2%, 51.8%, and 81.0% for Fig. 1 and 5.3%, 17.6%, 40.8%, 72.4%, and 91.3% in Fig. 2. This clearly shows the degradation of the extrapolation with distance from the patch. Within the patch itself, the error was 0.4 and 0.6%, respectively.

In the second set of figures the results for a higher frequency, 5100 Hz, are shown. This frequency is close to the coincidence frequency of the plate (the latter being 6267 Hz for an infinite plate). Again, Figs. 3 and 4 show two 11×11 patches located in different positions above the plate (the

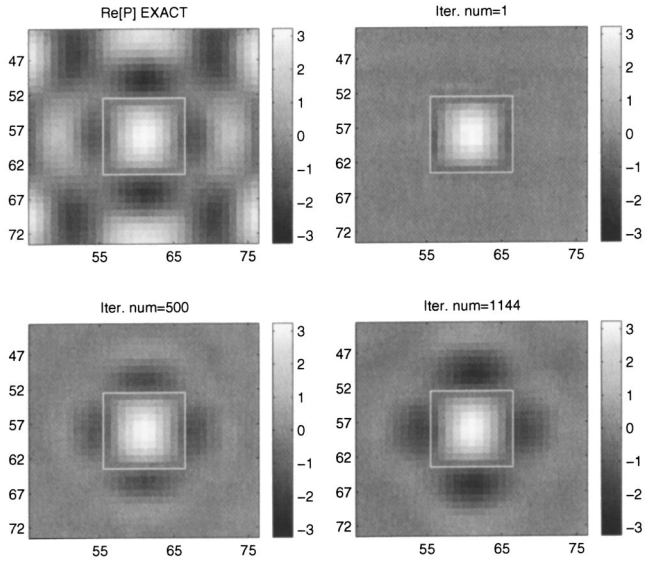


FIG. 2. The same case (2100 Hz) as Fig. 1, except the patch is shifted to a different location.

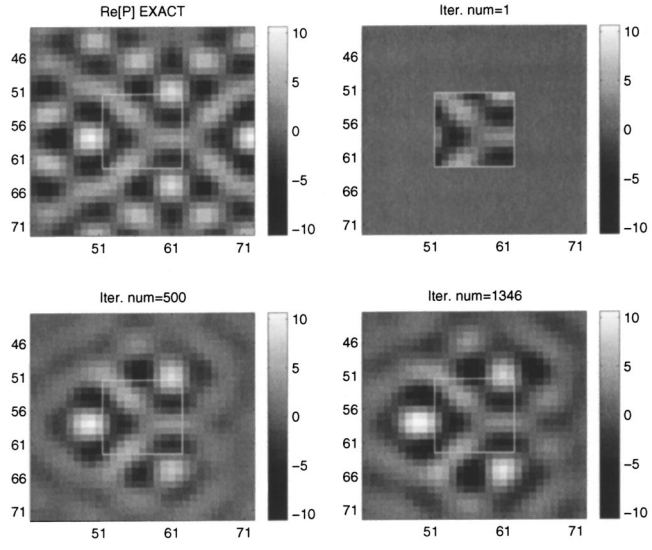


FIG. 3. The same case as Fig. 1, except the drive frequency is now 5100 Hz.

same positions as in Figs. 1 and 2, respectively) and a full aperture of 51×51 points. It is quite remarkable how accurately the pressure field is extrapolated near the patch and that as one moves farther away the phase of the pressure field is retained, even if the amplitudes are small. In this example the L_2 errors for these two cases were 6.4%, 20.3%, 40.7%, 66.9%, and 87.1% and 6.7%, 19.5%, 51.0%, 127.5%, and 114.5%, respectively.

One final important comment on the procedure above. The aperture (51×51) was quite large compared with the patch size (11×11). The reason for the large aperture compared to the size of the data patch is that in our studies the iteration equations were implemented using a DFT approach instead of the SVD since the geometry was planar. The DFT approach is the classic NAH approach for this geometry, and has the advantage of speed over the SVD. The large aperture guarantees that the replicated sources resulting from a DFT approximation of the continuous Fourier transform[1] are far enough apart that their radiation into the extrapolation region

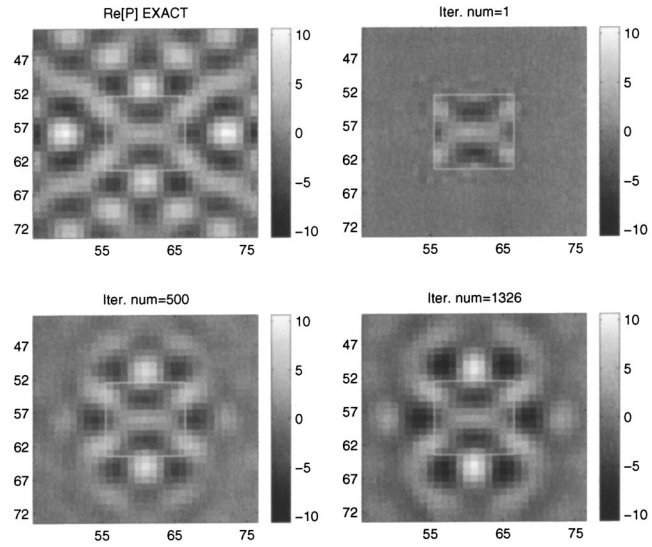


FIG. 4. The same as Fig. 2, except the drive frequency is now 5100 Hz.

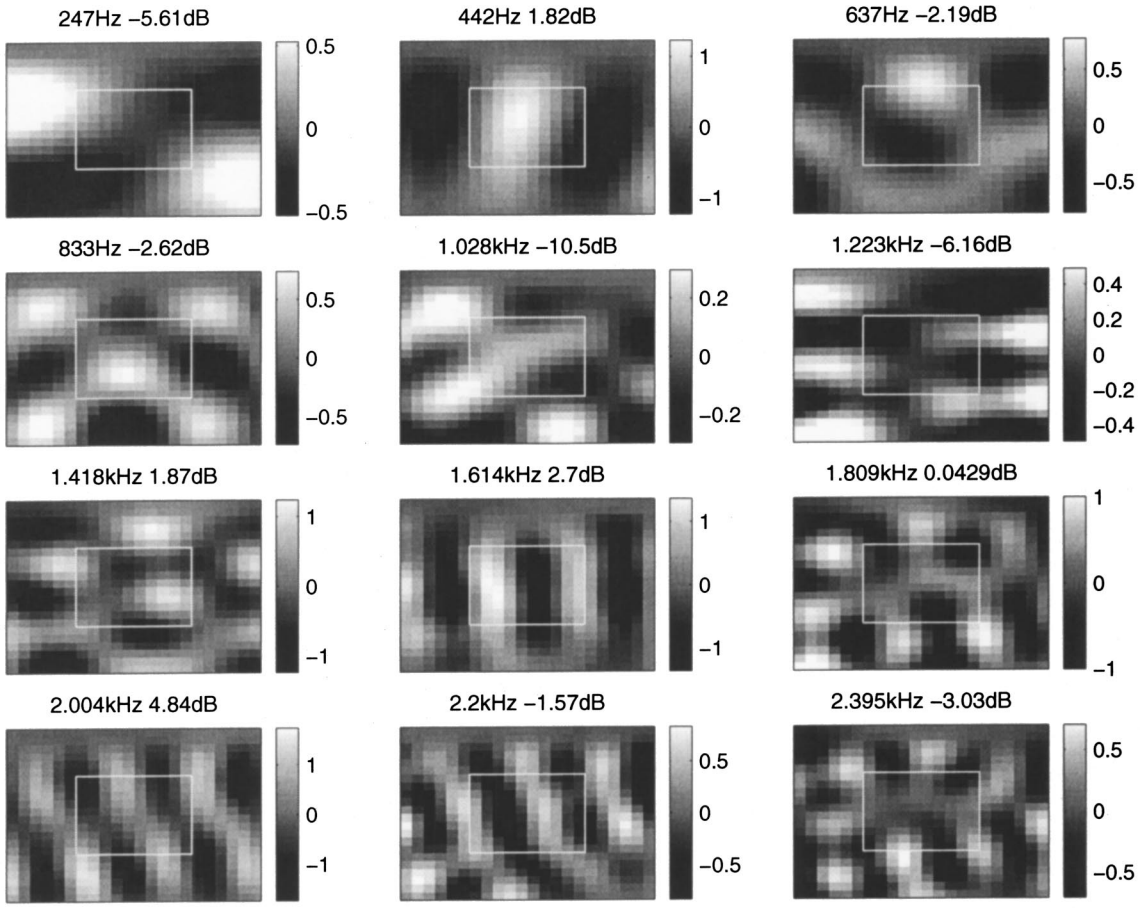


FIG. 5. Experimental data, actual measured pressure (per unit force) at 12 frequencies is shown for comparison with the continued field in Fig. 6. The 10×10 patch that was selected to test the iteration algorithm is indicated by the white box.

is much less than the actual field there. (It is interesting to note that the SVD does not suffer from this replicated source problem.) Furthermore, the closed form expression of the transfer function **G** is known in the DFT approach so one can easily construct and analyze the needed filters. Note that Eq. (3) and Eq. (4) do not change in form when using the DFT instead of the SVD, although **G** is now a complex-valued diagonal matrix and its ordering is determined by the radius in k space. **U** and **V** can be constructed from the one-dimensional DFT,[7] but this is not necessary in practice since one only needs the Fourier coefficients, which are generated by the matrix-vector multiply $\mathbf{U}^H\hat{p}^{\delta}$. These Fourier coefficients are determined directly from the two-dimensional DFT of $\hat{p}^{\delta}$.

V. EXPERIMENTAL EXAMPLE: POINT-DRIVEN SIMPLY SUPPORTED PLATE

A steel rectangular plate of dimensions 55.9×26.7 cm and thickness of 1.6 mm was driven by a Wilcoxon F3/F9 inertial shaker located near a corner of the plate. The edges of the plate were fixed to simulate approximately a simply-supported boundary condition. The resulting applied force was used to normalize the pressure measured with an ACO 7046 1/2 in. microphone attached to an automated $x-y$ scanner. The pressure was measured at a fixed distance of 0.4 cm from the plate surface, on a grid of points covering the complete plate surface. The grid step size in both the x and y directions was 1.25 cm. In order to exercise the algorithm discussed above, we extracted a small patch of data from this measured pressure consisting of 10×10 points, located at the center and corresponding to a square area 12.5 cm on a side.

The objective in the experimental data study was to use the 10×10-point pressure patch, using the DFT continuation algorithm to continue this data into a 20-point rim around the patch (in the same plane), and then compare the continued pressure with the measured pressure in this added rim. We present the results for 12 equally spaced frequencies chosen from the complete dataset (consisting of 1280 frequencies). Figure 5 presents the real part of the actual measured pressure located in a 22×22 point area, which includes the region chosen for the 10×10 patch, the latter indicated by a white outline. The frequencies are written above each of the 12 mosaics. The number next to the frequency is the maximum level in decibels *re* one Pascal per Newton force. Figure 6 shows the real part of the ''analytically'' continued field using the DFT iteration algorithm introduced above for the same region as shown in Fig. 5 and plotted with the same color scales for each frequency. Note that when using the iteration algorithm the six-point area outside the white box shown in Fig. 6 is initially filled with zeros. A comparison of the two figures shows that the continued field is close to the actual measurement and that the agreement degrades as one moves farther outside the boundaries of the patch, consistent

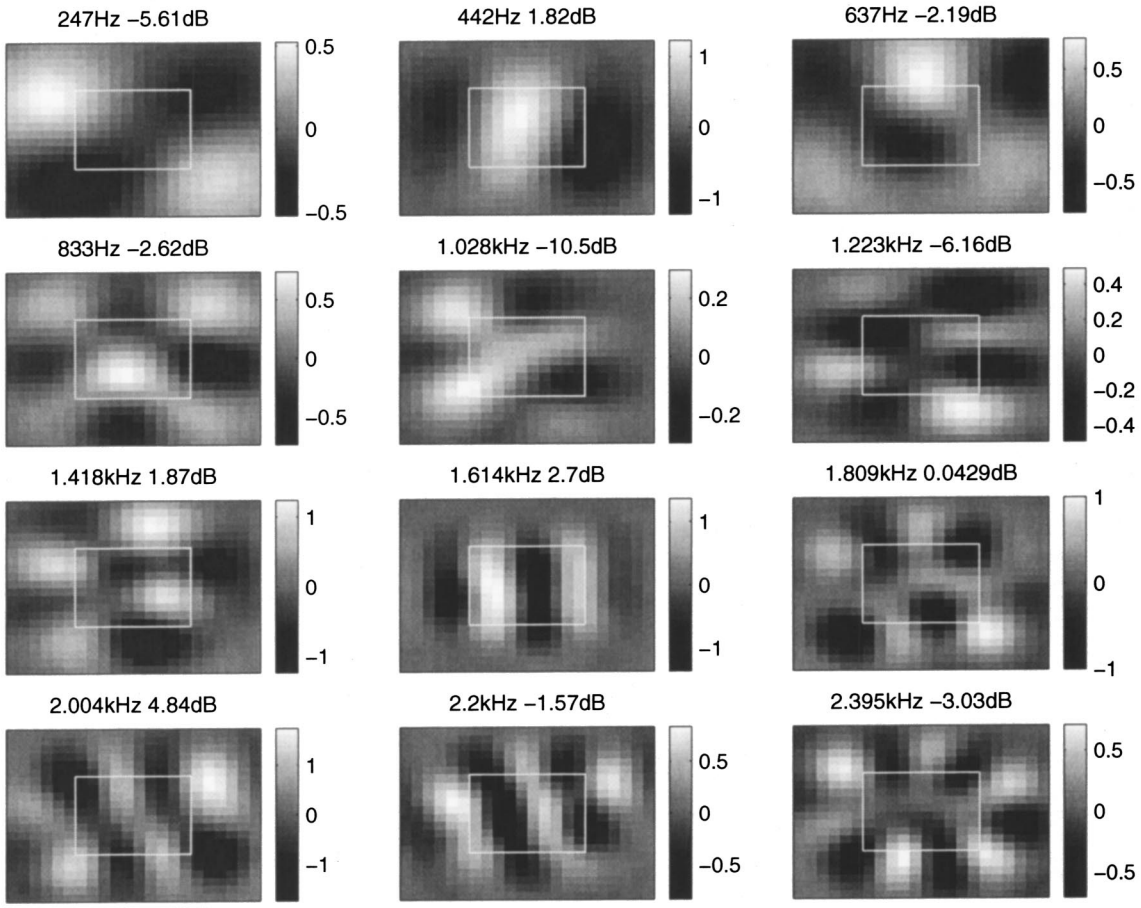


FIG. 6. Continued pressure (per unit force) derived from the iteration algorithm using the DFT approach. The same region of pressure is shown as displayed in Fig. 5. The area outside the white box is the region of ''analytic'' continuation.

with the results presented for the numerical case in the last section. Also evident in Fig. 6 is the taper of the pressure field as one moves away from the patch that resulted from the iteration algorithm.

Finally, to emphasize that the iteration procedure can be implemented successfully using the SVD [see Eq. (4)] instead of the DFT formulation, we include Fig. 7 applied to the same patch hologram shown within the white outline in Fig. 5. In this case only a 6-point rim was used instead of the 20-point rim used for the DFT. A comparison of Fig. 6 and Fig. 7 shows that the results of the DFT and SVD approaches are almost identical. This result is to be expected if one accepts the assumption that the continued field is unique (if one disregards the taper). Furthermore, since the transfer function leading to Eq. (4) can be written for any geometry, given a boundary element discretization of the vibrating surface, then the iteration scheme based on the SVD can be applied without modification for nonplanar geometries.

With regard to the application of the cutoff rule, Eq. (13), we required a minimum of 450 iterations before applying the cutoff rule that was almost always sufficient for convergence. In this case an *ad hoc* factor of 0.2 was used instead of 0.1 in Eq. (13).

VI. CONCLUSIONS

It should not be surprising that pressure field extrapolation accomplished using an iteration procedure is unable to reproduce the far field from the given data and instead produces a field that tends to zero there. The radiation model based on the Rayleigh integral for planar geometries, or the Helmholtz integral equation for general geometries, is a very local one due to the closeness of the conformal surfaces. In other words, information in the transfer function **H** relating points in the measurement aperture to points far outside of it is too weak to overcome the inversion perturbations. One can attribute this localness to the $1/R$ decay of the Green's function and a separation distance between the two surfaces equal (or nearly equal) to the lattice spacing, in our examples. This localness was more evident in efforts (not reported on here) to develop an extrapolation algorithm using Rayleigh's second integral, which relates *pressure* on the vibrating surface to pressure on the measurement surface. In this case the extrapolation of the measured pressure, which resulted dropped to zero more quickly outside the measurement aperture that the results based on Rayleigh's first integral discussed in this paper, implying an even more local relationship. This seems to be consistent with the fact that the Green's function drops off as $1/R^2$ as one moves away from the boundary for Rayleigh's second integral. On the other hand, this local relationship would seem to imply that as we move the two surfaces farther apart our extrapolation would succeed at larger distances from the edges. This may be true, and is a fertile area for more research, but was not of interest here since larger separation distances are generally not practical for NAH.

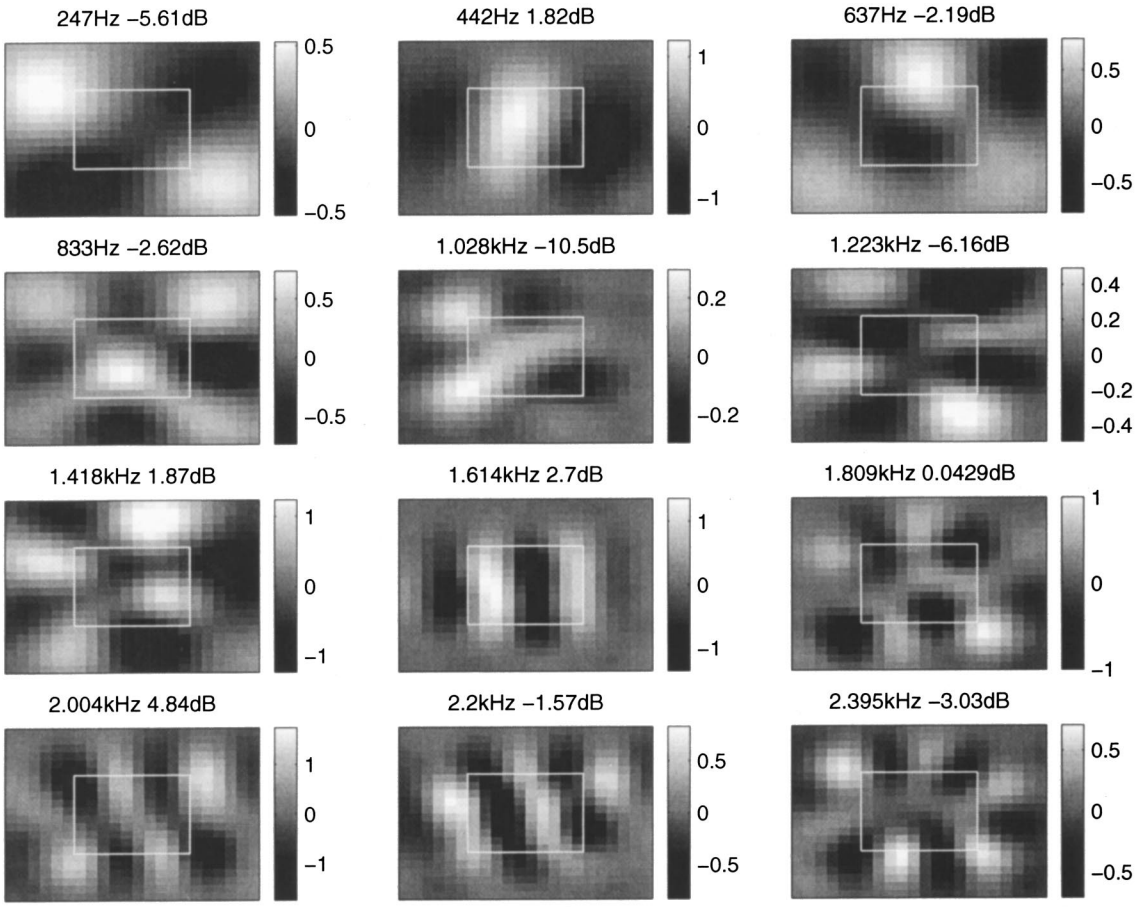


FIG. 7. The same as Fig. 6, except the SVD was used instead of the DFT in the iteration algorithm. The results are almost identical.

An interesting question arises: given the field in Ω_i is the field in Ω unique? If there only one acoustic field in Ω_0 that satisfies the wave equation and the radiation condition and that matches $p(x)$, $x \in \Omega_i$? We have no proof of uniqueness, although we hopefully can make uniqueness seem plausible by the following. If the acoustic field is known in a small region (what the mathematicians call a ball) that contains part of the measurement surface, then the *unique continuation principle* (Ref. 21, pp. 179–180) guarantees uniqueness of the field outside of the ball, in particular, on the sheet that is the tangential extension of the hologram patch. In other words, there is one and only one field in all space that is not null and that matches the prescribed field in the ball. Of course, the hologram is only two-dimensional and is not a ball, but given the additional information that the source location is known, it seems plausible that this principle applies and thus there is one and only one field corresponding to the continuation of the measured pressure field. We have tried to demonstrate here that the iteration algorithm reproduces the unique field close to the boundaries of the measurement patch.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. Experimental data was provided by Brian H. Houston and Peter C. Herdic at the Naval Research Laboratory.

[1] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, London, UK, 1999).
[2] R. F. Millar, "The Rayleigh hypothesis and a related least-squares solution to scattering problems for periodic surfaces and other scatterers," Radio Sci. **8**, 785–796 (1973).
[3] P. M. Van Den Berg and Jacob T. Fokkema, "The Rayleigh hypothesis in the theory of diffraction by a cylindrical obstacle," IEEE Trans. Antennas Propag. **27**, 577–583 (1979).
[4] R. F. Millar, "The location of singularities of two-dimensional harmonic functions. I: Theory," SIAM (Soc. Ind. Appl. Math.) J. Math. Anal. **1**, 333–343 (1970).
[5] D. Maystre and M. Cadihac, "Singularities of the continuation of fields and validity of Rayleigh's hypothesis," J. Math. Phys. **26**, 2201–2204 (1985).
[6] B.-Ki Kim and J.-Guon Ih, "Design of an optimal wave-vector filter for enhancing the resolution of reconstructed source field by NAH," J. Acoust. Soc. Am. **107**, 3289–3297 (2000).
[7] E. G. Williams, "Regularization methods for near-field acoustical holography," J. Acoust. Soc. Am. **110**, 1976–1988 (2001).
[8] M. R. Bai, "Acosutical source characterization by using recursive Wiener filtering," J. Acoust. Soc. Am. **97**, 2657–2663 (1995).
[9] E. G. Williams, "Numerical evaluation of the radiation from unbaffled finite plates using the FFT," J. Acoust. Soc. Am. **74**, 343–347 (1983).
[10] K. Saijyou and S. Yoshikawa, "Reduction methods of the reconstruction error for large-scale implementation of near-field acoustical holography," J. Acoust. Soc. Am. **110**, 2007–2023 (2001).
[11] M. R. Bai, "Application of bem (boundary element method)-based acoustic holography to radiation analysis and sound sources with arbitrarily shaped geometries," J. Acoust. Soc. Am. **92**, 533–549 (1992).
[12] S.-Chon Kang and J.-Guon Ih, "Use of nonsingular boundary integral formulation for reducing errors due to near-field measurements in the boundary element method based near-field acoustic holography," J. Acoust. Soc. Am. **109**, 1320–1328 (2001).
[13] Z. Zhang, N. Vlahopoulos, T. Allen, and K. Y. Zhang, "A source recon-

struction process based on an indirect variational boundary element formulation,'' Eng. Anal. Boundary Elem. **25**, 93–114 (2001).
[14] W. A. Veronesi and J. D. Maynard, ''Digital holographic reconstruction of sources with arbitrarily shaped surface,'' J. Acoust. Soc. Am. **85**, 588–598 (1989).
[15] G. Borgiotti, A. Sarkissian, E. G. Williams, and L. Schuetz, ''Generalized near-field acoustic holography for axisymmetric geometries,'' J. Acoust. Soc. Am. **88**, 199–209 (1990).
[16] B-Ki Kim and J-Guon Ih, ''On the reconstruction of the vibro-acoustic field over the surface enclosing an interior space using the boundary element method,'' J. Acoust. Soc. Am. **100**, 3003–3016 (1996).
[17] P. A. Nelson and S. H. Yoon, ''Estimation of acoustic source strength by inverse methods. Part I: Conditioning of the inverse problem,'' J. Sound Vib. **233**, 643–668 (2000).
[18] E. G. Williams, B. H. Houston, P. C. Herdic, S. T. Raveendra, and B. Gardner, ''Interior nah in flight,'' J. Acoust. Soc. Am. **108**, 1451 (2000).
[19] Per C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems* (SIAM, Philadelphia, PA, 1998).
[20] V. A. Morozov, ''Choice of parameter for the solution of functional equations by the regularization method,'' Sov. Math. Dokl. **8**, 1000–1003 (1967).
[21] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems* (Springer-Verlag, New York, NY, 1996).

Oscillatory flow in jet pumps: Nonlinear effects and minor losses

A. Petculescu and L. A. Wilen[a)]
Department of Physics and Astronomy, Ohio University, Athens, Ohio 45701

(Received 15 May 2002; revised 14 November 2002; accepted 14 December 2002)

A nonresonant, lumped-element technique is used to investigate the behavior of tapered cylindrical flow constrictions (jet pumps) in the nonlinear oscillatory flow regime. The array of samples studied spans a wide range of inlet curvature radii and taper angles. By measuring the rectified steady pressure component developed across a jet pump as well as the acoustic impedance, the minor loss coefficients for flow into and out of the narrow end of the jet pump are determined. These coefficients are found to be relatively insensitive to all but the smallest curvature radii (i.e., sharp edges). For fixed radius of curvature, the inflow minor loss coefficient increases with increasing taper angle while the outflow coefficient remains relatively constant. For all of the samples, the steady flow minor loss coefficients are also measured and compared to their oscillatory flow counterparts. The agreement is good, confirming the so-called Iguchi hypothesis. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1543932]

PACS numbers: 43.20.Ye, 43.58.Bh, 43.35.Ud, 47.27.Cn [MFH]

I. INTRODUCTION

Consider steady, incompressible, laminar flow along a pipe containing a region of gradual decrease in cross section. Mass conservation requires that the mean velocity increases as the pipe narrows and energy conservation (i.e., the Bernoulli equation) dictates that the pressure decrease correspondingly. (Here, we concentrate on processes at the transition region; clearly the pressure will also decrease gradually along the pipe due to the usual viscous dissipation effects as in, e.g., Poiseuille flow). Now instead consider an *abrupt* decrease in the channel cross section. A sudden change brings about additional dissipative processes such as flow separation, vortex shedding, turbulence, and jetting. As a consequence, some kinetic energy is converted to heat and the pressure drop will be larger than in the gradual case. The resulting extra pressure drop is known as a minor loss. Minor losses occur at any kind of sharp transition or bend in a pipe and have been studied extensively for steady flow. Discussions of the phenomenon can be found in almost any fluid mechanics textbook.[1] Since the pressure drop is related to the Bernoulli effect, it is usually characterized by a dimensionless minor loss parameter K:

$$\Delta p_{\text{minor loss}} = \Delta p_{\text{observed}} - \Delta p_{\text{ideal}} = \tfrac{1}{2} K \rho u^2, \tag{1}$$

where ρ is the density and u is the average flow velocity in the narrowest cross-section region. $\Delta p_{\text{observed}}$ is the observed pressure drop (or rise) across a sudden transition in cross section. Δp_{ideal} is the pressure drop (or rise) across the transition predicted from Bernoulli's principle. For example, if the fluid velocity is uniform in cross section at the two locations where pressure is measured, then

$$\Delta p_{\text{ideal}} = \frac{\rho}{2}(u_2^2 - u_1^2), \tag{2}$$

where u_1 and u_2 are the velocities in regions before and after the transition, respectively. If the velocities are nonuniform, than the expression for the Bernoulli pressure change is more complicated. Experimentally, K is found to be constant for large enough Reynolds numbers. Minor loss coefficients have been measured and tabulated for a large variety of systems.[2] With some assumptions, K can also be calculated for the case of an abrupt expansion in cross section.[3] For example, for an abrupt expansion in a circular tube, and assuming uniform velocity profiles everywhere, K is given by the Borda–Carnot formula:

$$K = \left(1 - \frac{A_1}{A_2}\right)^2, \tag{3}$$

where A_1 and A_2 are the cross-sectional areas of the tube before and after the expansion, respectively.

In the current experiment, we will consider minor losses across *constrictions* in a uniform diameter tube. In this case, the ideal (Bernoulli) pressure drop may be neglected. We define the total minor loss coefficient for the constriction according to Eq. (1). In general, the value of K will depend on the direction of flow unless a constriction is symmetric.

When a constriction is subjected to high amplitude *oscillatory* flow, the situation is much more complex. The minor loss pressure drop translates into a nonlinear acoustic impedance. For an asymmetric constriction, a nonzero time-averaged pressure drop will develop. Additional length constants, such as the viscous penetration depth and the displacement amplitude, are introduced. Furthermore, the flow velocity varies over the course of the cycle and therefore all the behavior seen in steady flow (e.g., laminar, turbulent) over a range of velocities may appear.

The present work concerns the investigation of oscillatory minor loss effects in jet pumps, which are essentially asymmetric constrictions in a tube formed from two transitions in cross section connected by a conical taper. The name originates from common water jets that exit from the narrow end of a nozzle. A major motivation for the present work is the recent application of jet pumps by Backhaus and Swift[4] and Swift, Gardener, and Backhaus[5] to suppress Gedeon

a)Electronic mail: wilen@helios.phy.ohiou.edu

streaming[6] in a traveling wave thermoacoustic engine. The time-averaged flow produced by jet pumps placed in the engine was adjusted to cancel the Gedeon streaming which otherwise caused an unwanted thermal short between the heat exchangers. This application constituted an indirect study of jet pump properties and clearly indicated the need for a more detailed examination. Other possible interesting applications include the use of jet pumps in microscale flow applications[7] and the modeling of flow constrictions in the bloodstream.[8]

In this study, we investigate the nonlinear impedance and rectification properties of axisymmetric jet pumps of various shapes. Some of the issues that we address are the applicability of Iguchi's hypothesis (described in the following), the effect of background impedance on the jet pump behavior, and the role of flow separation.

II. LITERATURE REVIEW

Oscillatory minor losses have been studied in some detail for orifices and pipe bends. Sivian[9] appears to be the first to report that the impedance of an orifice depends on the oscillatory velocity amplitude. He found that the resistance increased approximately linearly with velocity at high amplitudes. The measurements, performed in air, went up to about 30 m/s. He gave a physical explanation for the results based on the Bernoulli equation.

Using an impedance tube technique, Ingard and Labate[10] measured the nonlinear impedance of orifices with a variety of dimensions at fairly low flow velocities (<10 m/s). They also imaged flow patterns directly using smoke and stroboscopic imaging. They found ''step-like'' wiggles in the dependence of the resistance on velocity that could be correlated to distinct transitions in the flow patterns. With these observations they were able to construct a phase diagram for different oscillatory flow regimes as a function of orifice dimensions, frequency, and velocity amplitude. They were also able to relate the orifice resistance to the rate at which acoustic energy was converted into the kinetic energy of an observed circulating flow.

Bies and Wilson[11] measured the impedance of a Helmholtz resonator employing an orifice up to velocity amplitudes of about 100 m/s. At high velocity amplitudes they found that the orifice resistance increased linearly with amplitude while the reactance was found to decrease monotonically. Ingard and Ising[12] found similar results using a hot wire technique to measure particle velocity directly up to amplitudes of about 40 m/s.

In a series of papers by Thurston and co-workers,[13–15] careful measurements of the nonlinear impedance of thin symmetric and asymmetric orifices were performed using water and other liquids. In some cases the flow patterns were also imaged using injection of ink. Higher harmonics in the pressure response were measured, as well as the time-averaged (or dc) pressure (for the asymmetric case). At high amplitudes, a linear dependence of resistance on velocity amplitude was found, along with a decrease in reactance. The amplitude dependence of the impedance, along with the production of harmonics (including the dc pressure for asymmetric orifices), could be explained using a quasisteady approach whereby the instantaneous behavior was given by the usual expression for minor losses in steady flow [Eq. (1)].

Iguchi and Ohmi[16] studied U-tube oscillations for fluids with an orifice placed across the bottom of the tube. From the frequency and decay time of the oscillations, they came to a conclusion similar to that of Thurston and co-workers; the results could be described by a quasisteady approach as stated earlier. Because Iguchi and co-workers[17–22] have performed extensive measurements showing that this quasisteady assumption is applicable in many situations involving high amplitude time-dependent flow, the assumption is sometimes referred to as ''the Iguchi hypothesis.''

Panton and Goldman[23] employed dimensional analysis to correlate measurements of orifice impedance from many of the above-cited sources. They considered various dimensionless variables and demonstrated that measurements of resistance and reactance could be unified if properly normalized and plotted against the appropriately normalized velocity amplitude. The results were used to argue that different conditions govern the onset of nonlinearities in the resistance and reactance.

A number of recent studies by Smith and co-workers have looked at the detailed structure of flows around orifices and tube ends.[24–26] These studies use powerful flow visualization and hot wire techniques to map out spatial and temporal velocity profiles. They provide very detailed information on the oscillatory and time-averaged flows and can be used to observe the onset of different turbulent regimes and test ideas of scaling. Combined with pressure measurements, such techniques can also measure impedances.

Olsen and Swift[27] studied minor losses in pipe bends using a U-tube oscillation technique similar to that of Iguchi and co-workers. They found an empirical expression that could fit most of their data but concluded that the quasisteady approach was not applicable for the range of parameters that they explored.

A description of work where a steady flow is superimposed on the oscillatory flow has been omitted from this brief review; while interesting, it is not directly relevant to the present discussion.

III. EXPERIMENTAL APPARATUS

We employ a straightforward nonresonant technique to measure acoustic impedances and time-averaged pressure drops. The apparatus, shown in Fig. 1, is the same one used to measure the complex density of single pores and porous materials.[28] A commercial midrange loudspeaker (acquired surplus, manufacturer unknown) drives one end of a bellows to generate oscillatory flow (and hence pressure) in the system. The other end of the bellows is sealed to a hole on one side of an aluminum block. A jet pump sample can be mounted and sealed with an O-ring to the opposite side. A gauge pressure sensor (Honeywell Micro Switch 24PCAFA2G), calibrated previously against an MKS Baratron (model 622A13TED), is mounted in the aluminum block. The Honeywell sensor is a Wheatstone Bridge whose arms are sensitive to the bending of a thin silicon diaphragm. The bridge is driven with a floating 10 V dc source. The oscillatory output pressure signal is demodulated with a vec-

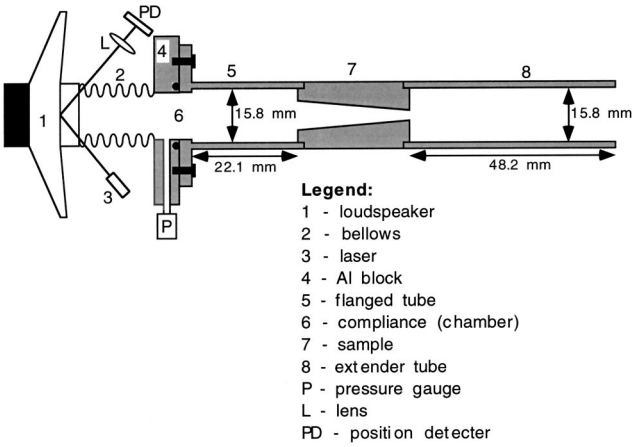


FIG. 1. Experimental apparatus for oscillatory measurements.

tor lock-in amplifier (Stanford Research SR 850). The bellows displacement is measured by a position detector that records the deflection of a laser beam reflected from a thin silicon mirror attached to the speaker cone. The oscillatory displacement signal is demodulated with a second vector lock-in amplifier (Stanford Research SR 830), which also provides the sinusoidal drive voltage (buffered by a power amplifier) for the speaker as well as the reference input for the other lock-in. Both lock-ins are interfaced to a computer. The time-averaged pressure signal is measured with a digital multimeter (Keithley 2000). The entire setup is enclosed in a Plexiglas™ box in order to reduce the effects of ambient pressure fluctuations. A typical set of data for one sample consists of the complex oscillatory pressure and displacement as well as the time-averaged pressure, as the speaker drive amplitude is swept from 0.05 to 1.50 V_{rms} at 96 Hz. These drives correspond to maximum particle velocities in the jet pump from 1.5 to 45 m/s and displacement amplitudes from 2 to 75 mm. Additionally, for calibration purposes, data are acquired with a blank flange, and also a cylindrical volume of known dimensions, mounted to the aluminum plate. Temperature, atmospheric pressure, and humidity are monitored during a run in order to account for changes in atmospheric density. The typical air density was 1.16 kg/m^3 $\pm$1%.

Steady flow measurements are performed in a separate apparatus, shown in Fig. 2. The jet pump is placed in series with a tube fed by a mass flow controller (Matheson 8272). Pure nitrogen flows through the system at a specified mass flow rate. The gauge pressure (Honeywell Micro Switch 24PCAFA2G) is measured upstream of the jet pump. Upstream of the pressure gauge, fiber wool and fine screens are placed in the tube to act as flow straighteners. A computer controls the flow rate and also reads the pressure. Data (pressure versus flow rate) are acquired over a specified range of flow, typically from 0 to 70 m/s (in the narrowest region). Temperature and atmospheric pressure are monitored during a run in order to calculate nitrogen density. The typical nitrogen density was 1.13 kg/m^3 $\pm$1%.

IV. SAMPLES

The jet pump samples are aluminum cylinders with cone-shaped holes (Table I). The holes were machined using

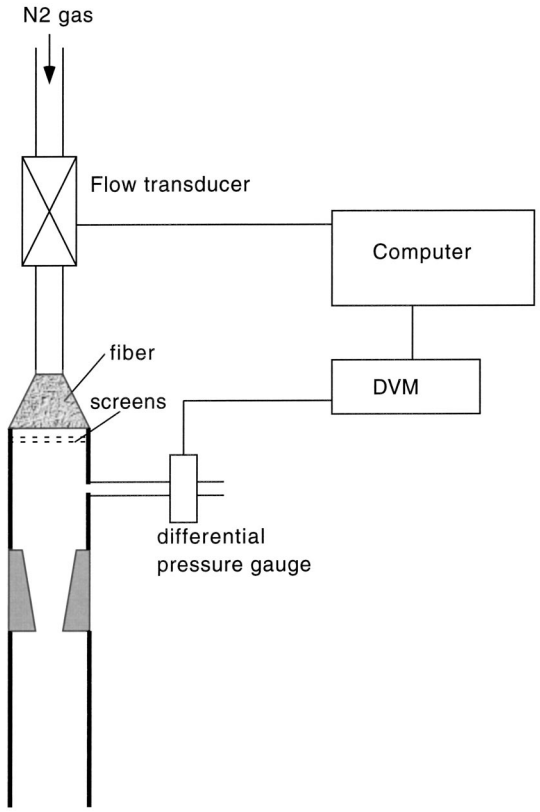
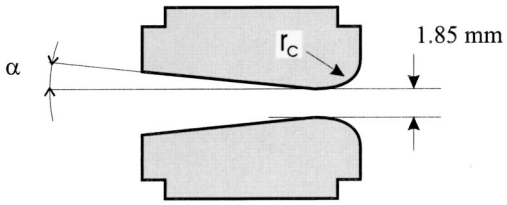


FIG. 2. Experimental apparatus for steady flow measurements.

a CAD-driven CNC milling machine. The edges at the narrow end of the conical hole were rounded with a specified radius of curvature. The edges at the wide end were left sharp. For all samples, the nominal waist (narrowest point in the cross section) diameter is 1.85 mm with variations within 1.0%. The waists were measured with a resolution of 0.0025 mm using stainless steel high-precision gage pins.[28] For all of the samples, the distance from the waist to the wide end was equal to 25.4 mm. Hence, the overall length of the

TABLE I.



curvature radius [mm] \ cone half-angle [degrees]	3	5	7	10
0	X		X	
0.11	X		X	
0.23	X		X	
0.46	X		X	
0.89	X		X	
1.40	X		X	
1.85	X	X	X	X

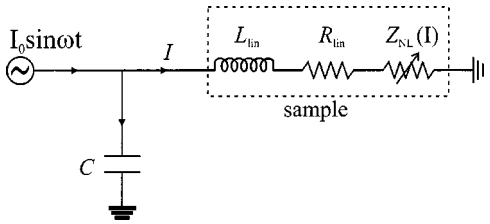


FIG. 3. Electrical analog of the acoustic circuit.

samples varied slightly, depending on the radius of curvature of the rounded edges. For the ''zero'' curvature samples, the edges were left sharp; clearly the actual curvature is finite but small. Four additional samples were made with a ''zero'' degree angle (i.e., straight circular tubes) and were rounded at both ends with a radius of curvature equal to 1.85 mm. For reference, the ratio of the viscous penetration depth to the waist diameter is 0.12 at 96 Hz in air. The Reynolds number ranged from 0 to 7000 for the dc measurements and from about 0 to 4000 for the ac measurements (calculated from the velocity amplitude at the waist and the waist diameter). The diameter of the tube into which the jet pump is inserted is 15.8 mm.

V. DATA ANALYSIS

The method of data analysis used to calculate sample impedances is almost identical to that described in earlier work.[28] A brief description is included here for completeness. We use a lumped-element approach in analyzing the data since the length of the system (approximately 15 cm) is small compared to an acoustic wavelength (approximately 3.6 m). Figure 3 shows a simple electrical analog of the acoustic circuit. From the measured complex pressure and displacement we can calculate the input impedance to the system, which comprises the wide compliant region acoustically in parallel with the jet pump. The compliance can be measured separately by using a sample plugged at the end near the compliant volume. A straightforward subtraction then yields the impedance of the jet pump alone,

$$(1/Z)_{\text{sample}} = (1/Z)_{\text{open}} - (1/Z)_{\text{plugged}}. \tag{4}$$

The resistance R and inertance L are defined as $\mathrm{Re}[Z]$ and $\mathrm{Im}[Z]/\omega$, respectively.

The actual measured impedance includes the impedance of the wide region between the pressure gauge and jet pump,[28] and the impedance of the wide extender tube, both in series with the jet pump. These impedances are primarily inertive (with a combined inertance of approximately 600 $\mathrm{kg/m^4}$). More importantly, they are amplitude independent as determined from direct measurements. This is reasonable because the velocity in this wide area is about 70 times smaller than the velocity in the jet pump waist (as determined from the area ratios) and minor losses depend on the velocity squared. In the present application we are mainly concerned with the nonlinear resistive part of the impedance; it will be seen that the linear impedance of the wide tube regions does not contribute to the minor loss coefficients in the analysis.

The system is calibrated with a cylindrical sample of well-characterized dimensions (and hence known impedance).[28] The oscillatory volume velocity in a jet pump sample is calculated from the measured oscillatory pressure and the jet pump impedance. The average velocity in the waist is taken to be the volume velocity divided by the waist area. Specific acoustic impedances (relative to the waist) may be calculated from the impedance and the waist area.

Figure 4(a) shows typical results for the resistance of a jet pump plotted against the average velocity amplitude u_0 in the waist (hereafter referred to simply as the velocity amplitude). The curve approaches a straight line at high amplitudes. The dependence of the resistance on amplitude results directly from the nonlinear minor losses. The best linear fit to the data at high amplitudes does not intercept zero, indicating that there is also a linear contribution to the resistance, probably associated with the interior region of the jet pump. Note also that the intercept from the high amplitude fit does not coincide with the low amplitude value of the resistance. This suggests that the linear flow resistance inside the jet pump is influenced by the inherently nonlinear flow conditions at the two transitions (i.e., at the ends of the jet pump). The inertance, shown in Fig. 4(b), is also amplitude dependent.

Figure 4(c) shows data for the time-averaged pressure, $\hat{p}$, plotted against the velocity amplitude u_0. The nonzero value for $\hat{p}$ is due to asymmetry in the flow. It is more instructive to plot time-averaged pressure divided by velocity, $\hat{p}/u_0$, (having the dimensions of specific resistance) versus u_0, shown in Fig. 4(d). Here also, the curve approaches a straight line with nonzero intercept at high amplitudes. In this case, the nonzero intercept implies that the (high amplitude) linear pressure drop in the jet pump depends on the direction of flow, again suggesting that the linear flow resistance inside the jet pump is influenced by the flow conditions at the transitions.

With two simplifying assumptions, the high amplitude slopes of the above-mentioned curves can be used directly to calculate high Reynolds number minor loss coefficients for oscillatory flow. It turns out that one of these assumptions is not valid, and it will be necessary to revise the analysis. Nevertheless, it is instructive to discuss the simple line of reasoning first and then show how it must be modified.

Before beginning the discussion, it is important to specify exactly what minor loss coefficients are measured in the present experiment. Referring to Fig. 1, one notes that there are abrupt changes in cross section and hence minor loss contributions at both ends of a jet pump sample. The total minor losses can be written as follows:

$$\begin{aligned}
\Delta p_{\text{ml}} &= \tfrac{1}{2}\rho K_{1+} u_1^2 + \tfrac{1}{2}\rho K_{2-} u_2^2 \\
&= \frac{1}{2}\rho\left(K_{1+} + K_{2-}\frac{A_1^2}{A_2^2}\right) u_1^2, \quad u(t) > 0, \\
\Delta p_{\text{ml}} &= \tfrac{1}{2}\rho K_{1-} u_1^2 + \tfrac{1}{2}\rho K_{2+} u_2^2 \\
&= \frac{1}{2}\rho\left(K_{1-} + K_{2+}\frac{A_1^2}{A_2^2}\right) u_1^2, \quad u(t) < 0.
\end{aligned} \tag{5}$$

The subscripts ''1'' and ''2'' refer to the narrow and wide end of the jet pump, respectively, and positive and

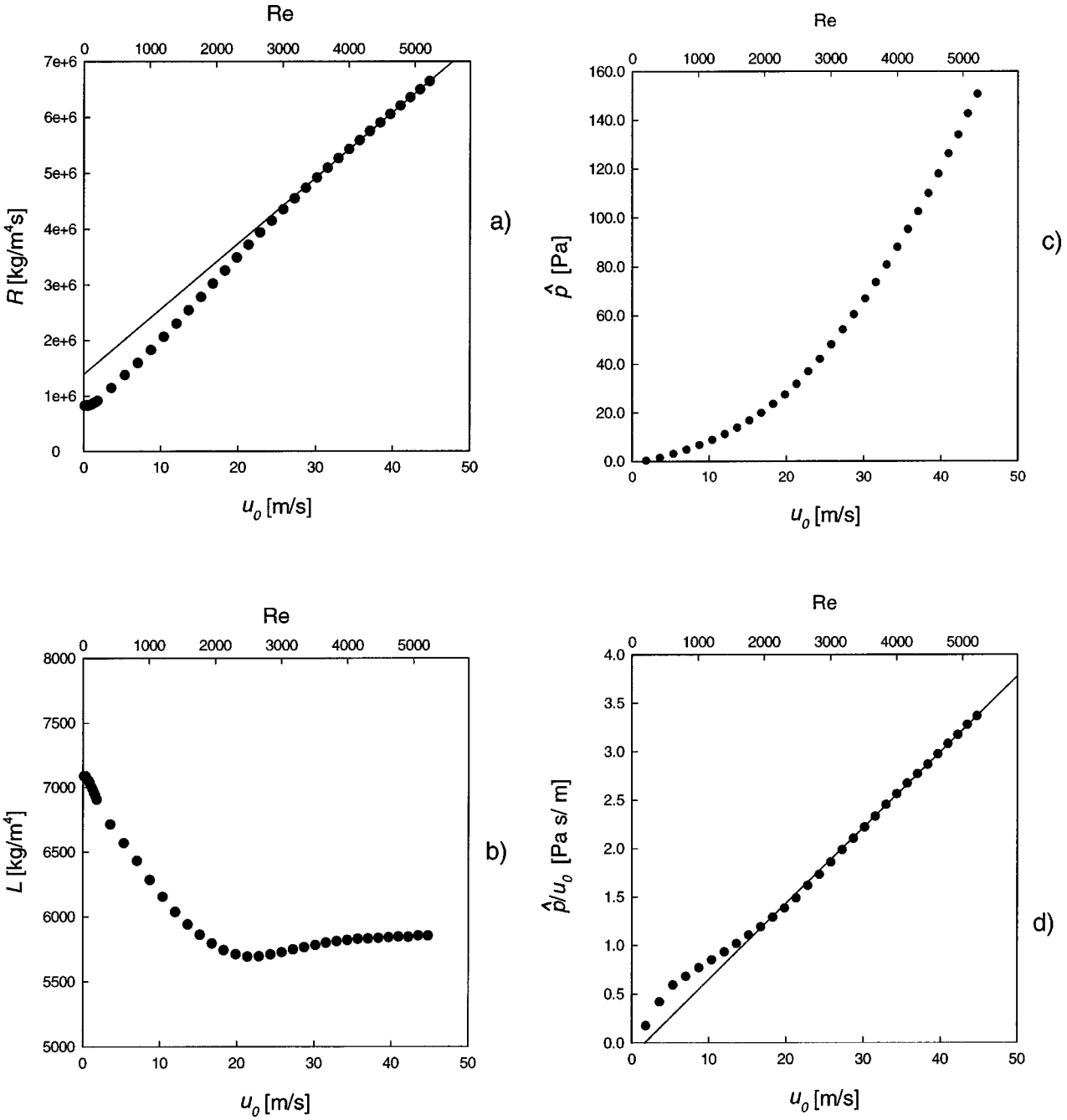


FIG. 4. (a) R, (b) L, (c) $\hat{p}$, and (d) $\hat{p}/u_0$ plotted against the velocity amplitude u_0 (bottom scale) and Reynolds number (top scale) for one jet pump (1.85 mm curvature, 7° angle). The solid lines in (a) and (d) are linear fits to the experimental data in the high amplitude region.

negative signs refer to flow out-of and in-to each end, respectively. By continuity, the velocity is smaller at the large end of the taper and a simple estimate shows that the minor loss contribution from the wide end (relative to the total) is below 4% for the 5°, 7°, and 10° samples, but may be as large as 15% for the 3° samples. Hence, the major contribution to the minor losses is from the narrow end and we define

$$K_{\text{out}}=K_{1+}+K_{2-}\frac{A_1^2}{A_2^2}, \quad K_{\text{in}}=K_{1-}+K_{2+}\frac{A_1^2}{A_2^2}, \tag{6}$$

where ''in'' and ''out'' refer to flow in-to and out-of the narrow end. For the straight pore samples, $A_1=A_2$, and we measure the quantity $K=K_{1+}+K_{2-}$ which by symmetry is equal to $K_{1-}+K_{2+}$. In the experiment, the actual pressure measured is the pressure inside the tube before the jet pump relative to the pressure outside the system. Technically, we should therefore take into account the pressure difference between the region inside the tube after the jet pump relative to the pressure outside the system (due to both minor losses and Bernoulli). However, this pressure difference is negligible compared to the minor loss pressure drop.

It is worth noting here that, because we measure the pressure drop across the whole jet pump (or straight pore) and not just across a single abrupt transition, we also see contributions that are linear in the velocity.

Thurston[13–15] and also Iguchi[16] have hypothesized that

for large amplitude oscillatory flow, minor losses can be accounted for by replacing steady flow quantities with their instantaneous counterparts. This hypothesis is motivated by the idea that for large enough amplitudes the flow has no memory of its past.[16–22] Specifically, one can write:

$$\Delta p_{\mathrm{ml}}(t)=\tfrac{1}{2}\rho K_{\mathrm{out}}u(t)^2, \quad u(t)>0,$$
$$\Delta p_{\mathrm{ml}}(t)=-\tfrac{1}{2}\rho K_{\mathrm{in}}u(t)^2, \quad u(t)<0, \tag{7}$$

where $u(t)$ is the instantaneous particle velocity at the narrow end, ρ is the density, and K_{out} and K_{in} are minor loss coefficients as defined earlier.

A second assumption is that the flow through the jet pump is assumed to be purely sinusoidal:

$$u(t)=u_0\sin(\omega t). \tag{8}$$

This is equivalent to saying that the jet pump is driven with a sinusoidal *current* source; hence the impedance of the jet pump does not affect the flow. Inserting Eq. (8) into Eq. (7), and applying Fourier analysis yields expressions for the amplitude-dependent resistance $\mathrm{R_{NL}}$ and time-averaged pressure $\hat{p}$:[4,15]

$$R_{\mathrm{NL}}=(2\rho/3\pi A)(K_{\mathrm{in}}+K_{\mathrm{out}})u_0,$$
$$\frac{\hat{p}}{u_0}=(\rho/8)(K_{\mathrm{out}}-K_{\mathrm{in}})u_0. \tag{9}$$

A is the area of the jet pump waist. These expressions predict constant slopes for the two experimental curves plotted in Figs. 4(a) and (d), and hence the measured slopes can be used to determine K_{in} and K_{out}. This procedure was used by Thurston *et al.*[13–15] to find minor loss coefficients for symmetric and asymmetric orifices. In Thurston's experiment, water was the fluid and the assumption that the flow was sinusoidal was valid because water is essentially incompressible. In the present experiment, however, air is used and the jet pump is in parallel with a compliant region. Therefore, the sinusoidal flow generated at the loudspeaker is split between the compliance and the jet pump. Hence, the flow through the jet pump will not be sinusoidal. Swift[29] has discussed this possibility in an exercise that compares results from an imposed sinusoidal pressure (''voltage source'') across the jet pump rather than an imposed sinusoidal flow (''current source''). The present case is intermediate, and additional calculations must be done to relate the experimental data to K_{in} and K_{out}. The procedure is as follows.

We find a numerical solution to the nonlinear lumped-element acoustic circuit equivalent to our system (Fig. 3). For convenience, we use electrical circuit notation where the voltage V and current I refer to the pressure and volume velocity, respectively. The nonlinear voltage across the jet pump can be written as follows:

$$V_{\mathrm{NL}}(I)=\frac{1}{4}\frac{\rho}{A^2}\{K_{\mathrm{out}}(|I|+I)I+K_{\mathrm{in}}(|I|-I)I\}, \tag{10}$$

where $|I|$ is the absolute value of I and A is the area of the waist in the jet pump. The total voltage across the jet pump is due to the nonlinear piece plus a linear resistance R_{lin} and inertance L_{lin}:

$$V=V_{\mathrm{NL}}(I)+R_{\mathrm{lin}}I+L_{\mathrm{lin}}\dot{I}. \tag{11}$$

The differential equation describing the system is given by

$$L_{\mathrm{lin}}C\ddot{I}+\left(R_{\mathrm{lin}}+\frac{dV_{\mathrm{NL}}}{dI}\right)C\dot{I}+I=I_0\sin(\omega t),$$
$$\frac{dV_{\mathrm{NL}}}{dI}=\frac{1}{2}\frac{\rho}{A^2}\{K_{\mathrm{out}}(|I|+I)+K_{\mathrm{in}}(|I|-I)\}. \tag{12}$$

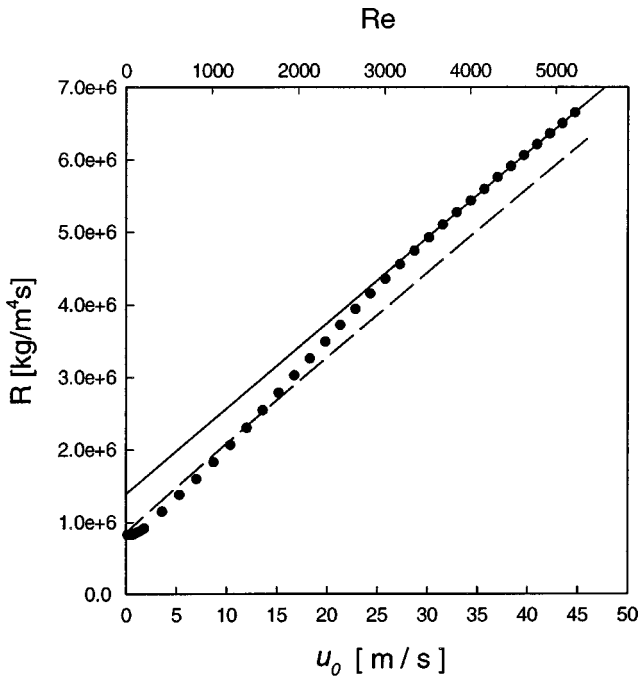
C is the compliance of the volume in front of the jet pump, I is the current through the jet pump, and I_0 is the amplitude of the current drive from the loudspeaker. C is determined from measurements with a jet pump sample plugged at the end closest to the bellows. Its value is approximately $9.55\times10^{-11}\,\mathrm{m^4\,s^2/kg}$ but varies slightly with atmospheric pressure. R_{lin} and L_{lin} are taken to be the values measured in the experiment at low drive amplitudes. K_{in} and K_{out} are adjustable parameters. The differential equation is solved on a one-dimensional time grid using a Verlet technique.[30] Standard checks are performed (comparing to analytical results in various limits, increasing the number of time steps, etc.) to ensure the validity of the procedure. Once a solution for $I(t)$ is found, the voltage across the jet pump is calculated from Eqs. (10) and (11). The time-averaged voltage and the in- and out-of-phase components of the fundamental voltage frequency response are found by numerical integration as follows:

$$\hat{V}=\frac{1}{\tau}\int_0^\tau V(t)dt,$$
$$V_{\text{in-phase}}=\frac{2}{\tau}\int_0^\tau V(t)\sin(\omega t)dt, \tag{13}$$
$$V_{\text{out-of-phase}}=\frac{2}{\tau}\int_0^\tau V(t)\cos(\omega t)dt,$$

where $\tau=2\pi/\omega$. The total impedance is given by

$$Z=\frac{V_{\text{in-phase}}}{I_0}+i\frac{V_{\text{out-of-phase}}}{I_0}.$$

The impedance for the jet pump alone, Z_{jet}, is then determined by subtracting off the compliance, just as is done in the experiment. The amplitude of the current through the jet pump is denoted by $I_{0\mathrm{jet}}$. From the results, curves of $R\equiv\mathrm{Re}[Z_{\mathrm{jet}}]$, $L\equiv\mathrm{Im}[Z_{\mathrm{jet}}]/\omega$ and $\hat{V}A/I_{0\mathrm{jet}}$ versus $I_{0\mathrm{jet}}/A$ can be generated and compared to the experimental curves for the same quantities. Figure 5 shows one such comparison. For each sample, K_{in} and K_{out} are adjusted until the numerical solution reproduces the experimental results in the high flow regime (>30 m/s, Re>3500). The slopes of the experimental curves can be reproduced exactly, but not the intercepts. This is because the linear resistance is slightly different (from the experimental low amplitude value) at high flow amplitudes, and is also different for inflow and outflow (as discussed earlier). It is possible to implement two linear resistance terms (for inflow and outflow) as inputs to the program and adjust them to achieve excellent agreement between the experiment and numerical solution (slopes and intercepts).



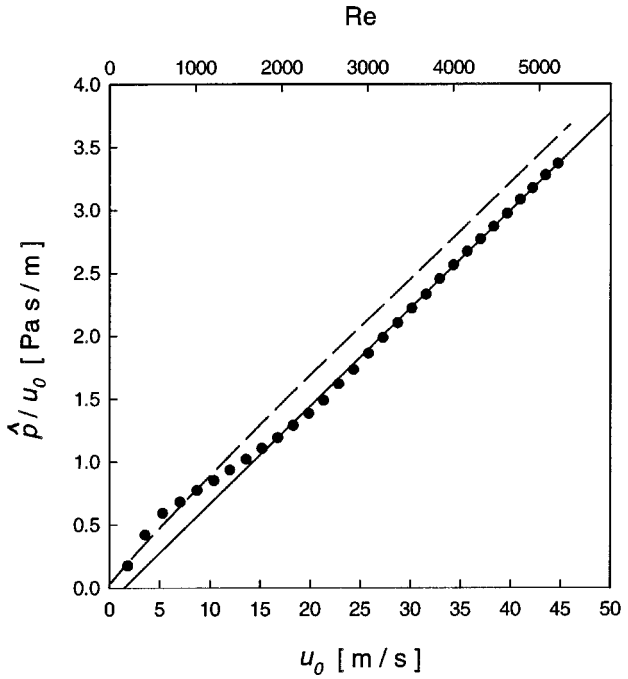


FIG. 5. Comparison between experimental data (black dots) and data generated from the numerical simulation (dashed lines) for the same jet pump as in Fig. 4. The solid lines are linear fits to the experimental data in the high amplitude region. K_{in} and K_{out} have been adjusted so that the slopes of the curves match at high amplitudes.

However, we checked that the slopes of the curves generated for R and $\hat{V}$ (and hence the results for K_{in} and K_{out}) are insensitive to such adjustments. The full dependence of L on amplitude is not captured in the model. However, again, the slopes of the curves generated for R and $\hat{V}$ are insensitive to changes in the input value for L_{lin} over a wide range. Using the numerical technique, values of K_{in} and K_{out} were determined for all of the jet pump samples and values of K were

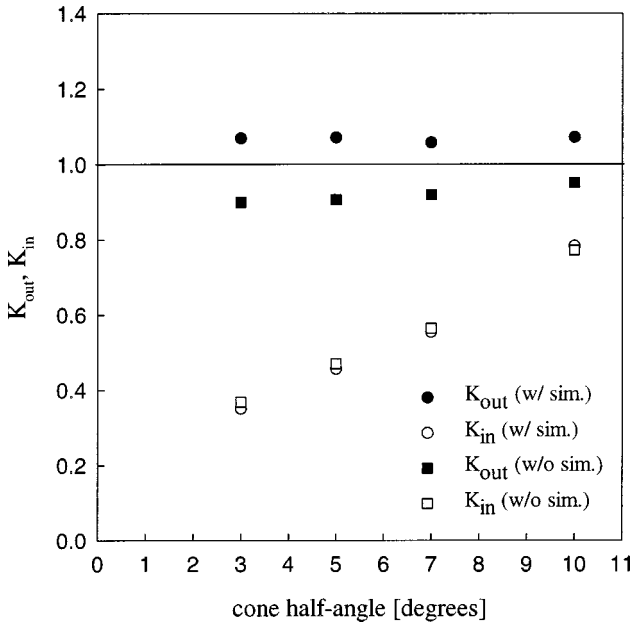


FIG. 6. ac results for K_{in} and K_{out} calculated using the full numerical procedure and also calculated from Eq. (9) combined with the slopes of the experimental curves.

determined for the straight pore samples. The observed (nonsinusoidal) shape of the pressure wave form for a given jet pump was qualitatively very similar to the wave form generated by the numerical solution, but we defer a full investigation of higher harmonics to a future study.

Figure 6 shows the oscillatory results for K_{in} and K_{out} for the largest curvature samples calculated using the full numerical procedure discussed earlier and also ''uncorrected values'' calculated directly from Eq. (9) combined with the slopes of the experimental curves. Note that the difference between corrected and uncorrected results is small for K_{in} but significant for K_{out}.

We also determined K_{in} and K_{out} using a model that assumed an imposed sinusoidal pressure on the jet pump (similar to that proposed in Swift[29] but including a dc pressure term which was adjusted to give zero time-averaged flow). Interestingly, the results were almost identical to the full numerical solution indicating that the jet pump acts as a high impedance with regard to its minor loss behavior. This model is described in the Appendix.

VI. dc FLOW MEASUREMENTS

Steady flow results for K_{in} and K_{out} were tabulated for many different geometries by Idelchik. The results for diffusers (equivalent to jet pumps for ''in-flow'') in Idelchik pertain to Reynolds numbers greater than 10^5 (evaluated for the waist diameter) whereas our measurements extend up to about 4000 and 7000 for the ac and dc cases, respectively. Hence, in order to make a direct comparison to oscillatory flow, and to check Iguchi's hypothesis, we performed steady flow measurements to determine K_{in} and K_{out}. Figure 7 shows typical results for p_{dc}/u vs u, where p_{dc} is the steady pressure across the jet pump and u is the average flow velocity in the narrow end. For each jet pump, data were acquired for inflow and outflow orientations.

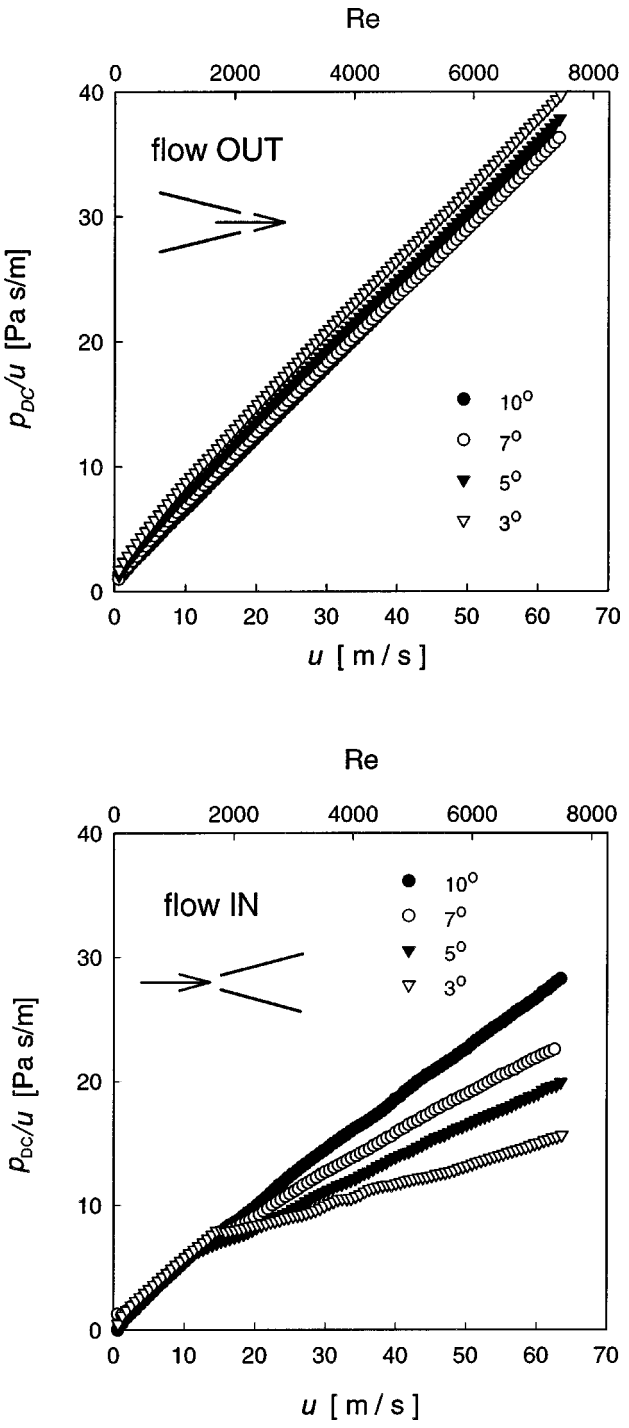


FIG. 7. Steady flow results for p_{dc}/u vs u (bottom scale) and Reynolds number (top scale) for the large curvature (1.85 mm) jet pumps.

Although the goal was to acquire numbers that could be compared with the ac results, the dc curves have intrinsic interest. For flow into the jet pump, there is a break in the curve at a well-defined velocity. Such behavior can be interpreted in terms of velocity dependent flow separation in the conical section of the jet pump. Also, for the outflow curves, the low amplitude data yield a minor loss coefficient close to 2, in agreement with theoretical predictions for abrupt expansion of laminar Poiseuille flow. We defer a detailed discussion of these features to a future publication.

For comparison with the ac results, we use the K values calculated from the high velocity (Re>5000) portion of the curves. We plot p_{dc}/u against u and take the slope to be equal to $1/2\rho K_x$ (x="in" or "out" depending on the orientation of the jet pump.) Note that here, as for the ac results, the intercept is nonzero.

VII. SOURCES OF ERROR

Statistical errors are very small in all cases and result in error bars that are of order the size of the data points. There are a number of sources of systematic error. These include pressure gauge calibration errors, flow meter calibration errors, errors in measurements of dimensions, and errors in measurements of ambient pressure, temperature, and humidity. For the steady flow K_{in} measurements, there is an additional consideration that leads to uncertainties; because the curves are not perfectly linear in the high amplitude region, the slope depends somewhat on the range used for the linear fit. (This is also true for the ac curves, but to a smaller extent). An additional possible source of systematic error is that introduced by nonlinearity in the pressure gauge, which can produce a false time-averaged pressure signal.[31] We checked for this in the following way. We measured the time-averaged pressure for three jet pumps, with each jet pump mounted in the system in a "forward" and "reverse" orientation. (Forward is defined to be the case when the wide end of the jet pump faces the bellows/loudspeaker assembly.) Any contribution to the time-averaged signal from nonlinearity in the gauge will not reverse sign when the jet pump is reversed. We found that the time-averaged signal for the two orientations (for all three samples) were opposite in sign and equal in magnitude to within 5%. There are also uncertainties associated with using the simple lumped-element circuit model for the system, which neglects the compliance of the jet pump. It is possible to get a handle on these uncertainties by looking at the data with jet pumps reversed since the compliance of the jet pump appears in a different way in the circuit for the two orientations. Analyzing the data from the forward and reverse orientations, we found that the K values agreed to better than 0.02. Accounting for all the above-mentioned effects, we estimate that the total absolute error in the results for minor loss coefficients is ± 0.02 for the oscillatory measurements and ± 0.03 for the steady flow measurements.

There are also errors in the measured values for angle and curvature. The former are negligible, but the latter, of order 0.02 mm, are important for the smallest curvatures.

VIII. RESULTS AND DISCUSSION

The basic qualitative features of the ac results are consistent with analogous results from earlier orifice studies. Figure 8 shows results for R, L, and $\hat{p}/u_0$ for the large curvature jet pumps as a function of velocity amplitude. The resistance increases linearly with amplitude at high amplitudes and the inertance initially decreases with amplitude and then levels out (Fig. 8). For the orifice work, the decrease in inertance was attributed to an effective decrease in the mass participating in the oscillation. The basic idea is that some of the kinetic energy associated with the acoustic

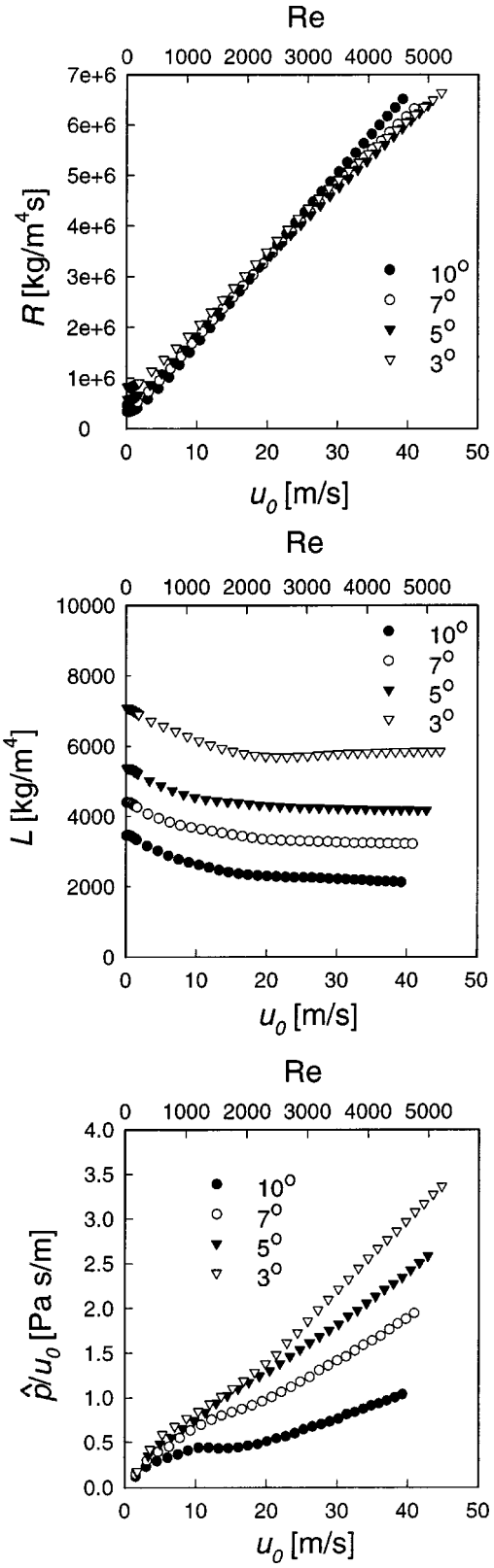


FIG. 8. Results for R, L, and $\hat{p}/u_0$ vs u_0 (bottom scale) and Reynolds number (top scale) for the large curvature (1.85 mm) jet pumps.

inertance of the orifice is converted to turbulent (jet-like) flow. Hence, the same loss in kinetic energy responsible for the pressure drop (leading to the nonlinear resistance) translates into a smaller acoustic mass.[10] It is likely that the explanation for jet pumps is similar.

Figures 9(a)–(c) show the results for K_{in} and K_{out} from both the oscillatory (ac) and steady flow (dc) measurements for all of the samples. Examination of these results shows reasonably good agreement between the ac and dc results, except for the sharpest inlets (smallest curvature).

The ac and dc results for K_{out} (for all samples) are close to 1.05. This is consistent with the idea that no flow separation occurs in the region of decreasing cross section. The minor losses are due to the abrupt expansion at the narrow end and are given theoretically by the Borda–Carnot formula [Eq. (3)], which predicts $K_{\text{out}}=1$. The Borda–Carnot formula assumes a uniform velocity distribution. A more realistic velocity distribution must decrease to zero at the wall, and hence values somewhat higher than 1 are expected.

Figure 9(a) shows results for K_{in} and K_{out} as a function of angle for the largest inlet curvature. The results for K_{in} are consistent with the idea that, as the angle is decreased, the point at which the flow separates from the wall occurs at larger diameters, leading to reduced minor losses. The results for K_{in} are compared with a formula for stepped diffusers found in Idelchik (p. 265, No. 73).[2] The geometry in Idelchik is apparently the same as the one used in the present experiment but, as noted earlier, Idelchik's results pertain to Re $>2\times10^5$. The trend of decreasing K_{in} with angle is similar although the exact values are not in good agreement.

Figures 9(b) and (c) show results for K_{in} and K_{out} as a function of curvature for two different angles. K_{in} is relatively constant for curvatures above 0.23 mm. The significant increase of K_{in} at the smallest curvature is likely due to the formation of a vena contracta just inside the inlet, similar to what occurs for a sharp-edged orifice geometry.[1] The small curvature behavior is somewhat different for the ac and dc measurements, but the qualitative trend is the same. The dip in K_{in} for the 3° sample at a curvature of 0.1 mm is somewhat puzzling and we do not have a simple physical explanation. Nevertheless, the feature is reproducible and is seen in both the ac and dc data.

A useful check on the validity of the experimental procedure is provided by measurements on straight pore samples with rounded openings. As discussed earlier, these measurements determine the minor loss coefficient $K=K_{1+}+K_{2-}$. Theoretically, one would expect a value close to (and larger than) 1 for K_{1+}, as discussed earlier, and a value close to 0 for K_{2-}, since the rounded entrance tends to prevent the formation of a vena contracta and flow separation is not expected to occur in the straight pore. One must be careful in these measurements, however, because the Reynolds number is large enough that turbulence in the middle section of the tube may also contribute to a nonlinear resistance. To disentangle these effects, we performed ac and dc measurements on straight tubes of four different lengths. Figure 10 shows the results for $K_{1+}+K_{2-}$ plotted against sample length. Clearly there is a nonlinear contribution to the pressure drop from the middle of the tube. The ac and dc results are in rough agreement and both extrapolate back to a value close to 1 for zero length. From these measurements we can conclude that some contribution to the minor losses measured in the jet pumps may also come from the interior of the (conical) tube although to a much smaller extent (compared to the straight pores) because the average diameter is larger.

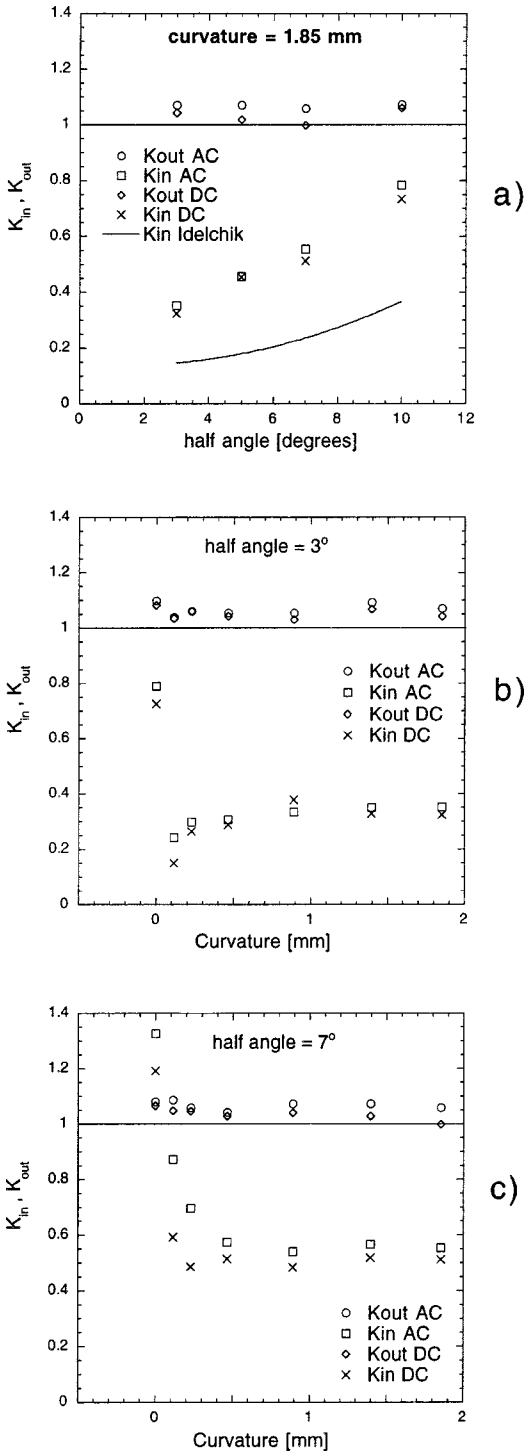


FIG. 9. Results for K_{in} and K_{out} from both the oscillatory (ac) and steady flow (dc) measurements. (a) K_{in} and K_{out} vs angle for 1.85 mm curvature jet pumps. Solid curve is from Idelchik. (b) K_{in} and K_{out} vs curvature for 3° jet pumps. (c) K_{in} and K_{out} vs curvature for 7° jet pumps.

Further investigation of straight pore minor losses and interior losses will be deferred to a future publication.

IX. CONCLUSIONS

The main conclusion of this work is the fairly good agreement between the ac and dc results for K_{in} and K_{out}, supporting Iguchi's quasisteady hypothesis. To resolve this agreement, it was important to pay attention to the impedance of the jet pump relative to that of the system. In our arrangement, the jet pump behaved as if it had a high impedance. If jet pumps are to be employed in thermoacoustic devices to prevent streaming, some attention must be given to this point. A modification in the system configuration resulting in a different impedance at the jet pump location will require some extra retuning. A more detailed test of Iguchi's hypothesis could be performed by using the explicit curves (p_{dc} vs u) measured for steady flow to see if they can predict ac results for R and $\hat{p}$ vs u_0. This will be the topic of future work.

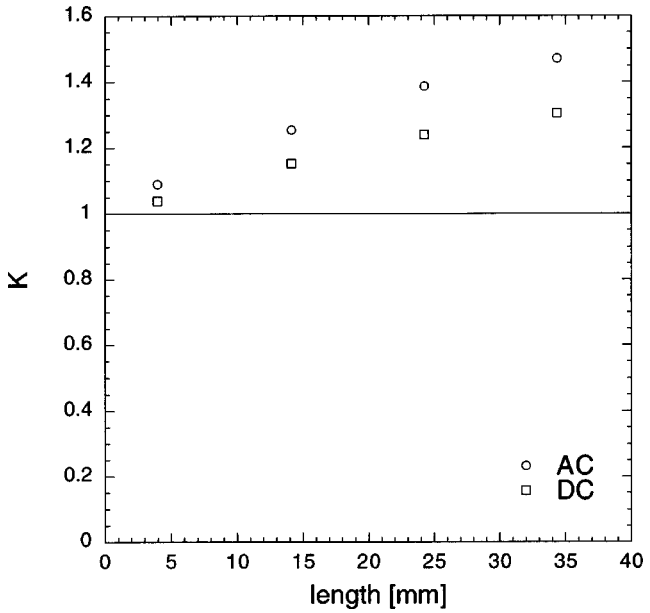


FIG. 10. Results for $K=K_{1+}+K_{2-}$ for straight rounded tubes plotted against sample length.

Swift, Gardener, and Backhaus,[5] and Backhaus and Swift[4] tested many of the above-mentioned ideas indirectly by employing jet pumps with different inlet conditions or different angles to suppress streaming in acoustic traveling-wave devices. There is good qualitative agreement between their conclusions and those reached here. To achieve optimum performance in new traveling-wave devices, it may prove useful to test and improve jet pump designs with a setup similar to that described here.

ACKNOWLEDGMENTS

The authors would like to thank Anthony Atchley, Barton Smith, and Ray Wakeland for very fruitful discussions. Work supported by the Office of Naval Research.

APPENDIX

In this section we describe briefly the voltage source model we alluded to at the end of Sec. V. Here, again, we use the electrical circuit notation. We assume that a voltage of the following form is imposed across the jet pump:

$$V(t)=V_0 \sin \omega t + V_{\text{dc}}. \tag{A1}$$

Following Eq. (7), the current response is given by

$$I(t) = +\sqrt{\frac{2A^2V_0}{\rho}}\sqrt{\frac{\sin\omega t+\epsilon}{K_{\text{in}}}}, \quad \sin\omega t+\epsilon>0,$$
$$I(t) = -\sqrt{\frac{2A^2V_0}{\rho}}\sqrt{\frac{-\sin\omega t-\epsilon}{K_{\text{out}}}}, \quad \sin\omega t+\epsilon<0, \tag{A2}$$

where $\epsilon = V_{\text{dc}}/V_0$.

For our system, we impose the condition that there is no time-averaged flow:

$$\int_0^{2\pi/\omega} I(t)\,dt = 0. \tag{A3}$$

Equation (A3) is solved numerically to determine ϵ in terms of K_{in} and K_{out}. With ϵ determined, the first Fourier coefficient of the in-phase current response is calculated (also numerically):

$$I_0 \equiv \frac{\omega}{\pi}\int_0^{2\pi/\omega} I(t)(\sin\omega t)\,dt = \alpha(\epsilon, K_{\text{in}}, K_{\text{out}})\sqrt{\frac{2A^2V_0}{\rho}}. \tag{A4}$$

The result is expressed in terms of a dimensionless constant α that depends only on ϵ, K_{in} and K_{out}. The linear resistance and the time-averaged voltage can now be written in terms of α and ϵ as follows:

$$\frac{V_{\text{dc}}}{I_0^2} = \frac{\rho}{2A^2}\frac{\epsilon}{\alpha^2}, \quad \frac{R}{I_0} = \frac{V_0}{I_0^2} = \frac{\rho}{2A^2}\frac{1}{\alpha^2}. \tag{A5}$$

The experimental slopes of V_{dc}/I_0 vs I_0 and R vs I_0 are used to determine α and ϵ, and then K_{in} and K_{out} are varied to match the measured values.

[1] R. W. Fox and A. T. McDonald, *Introduction to Fluid Mechanics* (Wiley, New York, 1992).
[2] I. E. Idelchik, *Handbook of Hydraulic Resistance*, 3rd ed. (Begell House, New York, 1996).
[3] W. M. Kays, "Loss coefficients for abrupt changes in flow cross section with low Reynolds number flow in single and multiple tube systems," Trans. ASME **72**, 1067–1074 (1950).
[4] S. Backhaus and G. W. Swift, "A thermoacoustic-Stirling heat engine: Detailed study," J. Acoust. Soc. Am. **107**, 3148–3166 (2000).
[5] G. W. Swift, D. L. Gardner, and S. Backhaus, "Acoustic recovery of lost power in pulse tube refrigerators," J. Acoust. Soc. Am. **105**, 711–724 (1999).
[6] D. Gedeon, "DC gas flows in Stirling and pulse-tube cryocoolers," in *Cryocoolers 9*, edited by R. G. Ross (Plenum, New York, 1997).
[7] R. L. Bayt and K. S. Breuer, "Viscous effects in supersonic MEMS-fabricated micronozzles," in Proceedings of the Third ASME Microfluid Symposium, Anaheim, CA, 1998.
[8] P. J. Kilner *et al.*, "Asymmetric redirection of flow through the heart," Nature (London) **404**, 759–761 (2000).
[9] L. J. Sivian, "Acoustic impedance of small orifices," J. Acoust. Soc. Am. **7**, 94–101 (1935).
[10] U. Ingård and S. Labate, "Acoustic circulation effects and the nonlinear impedance of orifices," J. Acoust. Soc. Am. **22**, 211–218 (1950); see also U. Ingård, "On the theory of acoustic resonators," *ibid.* **25**, 1037–1061 (1953).
[11] D. A. Bies and O. B. Wilson, Jr., "Acoustic impedance of a Helmholtz resonator at very high amplitude," J. Acoust. Soc. Am. **29**, 711–714 (1957).
[12] U. Ingård and H. Ising, "Acoustic nonlinearity of an orifice," J. Acoust. Soc. Am. **42**, 6–7 (1967).
[13] G. B. Thurston, "Nonlinear acoustic properties of orifices of varied shapes and edge conditions," J. Acoust. Soc. Am. **30**, 452–455 (1958).
[14] G. B. Thurston and C. E. Martin, "Periodic fluid flow through circular orifices," J. Acoust. Soc. Am. **25**, 26–31 (1953).
[15] G. B. Thurston, L. E. Hargrove, and B. D. Cook, "Nonlinear properties of circular orifices," J. Acoust. Soc. Am. **29**, 992–1001 (1957).
[16] M. Iguchi, M. Ohmi, and K. Maegawa, "Analysis of free oscillating flow in a U-shaped tube," Bull. JSME **25**, 1398–1405 (1982).
[17] M. Ohmi and M. Iguchi, "Critical Reynolds number in an oscillating pipe flow," Bull. JSME **25**, 165–172 (1982).
[18] M. Ohmi, M. Iguchi, and I. Urahata, "Transition to turbulence in a pulsatile pipe flow. 1 Wave forms and distribution of pulsatile velocities near transition region," Bull. JSME **25**, 182–189 (1982).
[19] M. Ohmi, M. Iguchi, K. Kakehashi, and T. Masuda, "Transition to turbulence and velocity distribution in an oscillating pipe flow," Bull. JSME **25**, 365–371 (1982).
[20] M. Ohmi, M. Iguchi, and I. Urahata, "Flow patterns and frictional losses in an oscillating pipe flow," Bull. JSME **25**, 536–543 (1982).
[21] M. Iguchi and M. Ohmi, "Transition to turbulence in a pulsatile pipe flow. 2. Characteristics of reversing flow accompanied by relaminarization," Bull. JSME **25**, 1529–1536 (1982).
[22] M. Iguchi and M. Ohmi, "Free oscillatory flow across a plate orifice in a U-shaped tube and the resistance law," Bull. JSME **25**, 1537–1543 (1982).
[23] R. L. Panton and A. L. Goldman, "Correlation of nonlinear orifice impedance," J. Acoust. Soc. Am. **60**, 1390–1396 (1977).
[24] B. L. Smith and A. Glezer, "Jet vectoring using synthetic jets," J. Fluid Mech. **458**, 1–34 (2002).
[25] B. L. Smith and A. Glezer, "The formation and evolution of synthetic jets," Phys. Fluids **10**, 2281–2297 (1998).
[26] B. L. Smith and G. W. Swift, "Synthetic jets at large Reynolds number and comparison to continuous jets," AIAA Pap. **2001-3030**, 1–19 (2001).
[27] J. R. Olson and G. W. Swift, "Energy dissipation in oscillating flow through straight and coiled pipes," J. Acoust. Soc. Am. **100**, 2123–2131 (1996).
[28] A. Petculescu and L. A. Wilen, "Lumped-element technique for the measurement of complex density," J. Acoust. Soc. Am. **110**, 1950–1957 (2001).
[29] G. W. Swift, *Thermoacoustics: A Unifying Perspective for Some Engines and Refrigerators* (Acoustical Society of America, New York, 2002).
[30] L. Verlet, "Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules," Phys. Rev. **159**, 98 (1967).
[31] B. L. Smith and G. W. Swift, "Measuring second-order time-average pressure," J. Acoust. Soc. Am. **110**, 717–723 (2001).

Effects of harmonic phase on nonlinear surface acoustic waves in the (111) plane of cubic crystals

R. E. Kumon[a] and M. F. Hamilton
Department of Mechanical Engineering, The University of Texas at Austin, Austin, Texas 78712-1063

(Received 11 February 2002; revised 5 October 2002; accepted 18 October 2002)

Spectral evolution equations are used to perform numerical studies of nonlinear surface acoustic waves in the (111) plane of several nonpiezoelectric cubic crystals. Nonlinearity matrix elements which describe the coupling of harmonic interactions are used to characterize velocity waveform distortion. In contrast to isotropic solids and the (001) plane of cubic crystals, the nonlinearity matrix elements usually cannot be written in a real-valued form. As a result, the harmonic components are not necessarily in phase, and dramatic variations in waveforms and propagation curves can be observed. Simulations are performed for initially monofrequency surface waves. In some directions the waveforms distort in a manner similar to nonlinear Rayleigh waves, while in other directions the velocity waveforms distort asymmetrically and the formation of shocks and cusped peaks is less distinct. In some cases, oscillations occur near the shocks and peaks because of phase differences between harmonics. A mathematical transformation based on the phase of the matrix elements is shown to provide a reasonable approximation of asymmetric waveform distortion in cases where the matrix elements have similar phase.
[DOI: 10.1121/1.1529170]

PACS numbers: 43.25.Fe, 43.35.Pt, 68.35.Iv, 62.65.+k [DEC]

I. INTRODUCTION

In a previous paper,[1] the authors used spectral evolution equations[2] to investigate the nonlinear propagation of surface acoustic waves (SAWs) in the (001) surface cut of nonpiezoelectric, cubic crystals. The present paper extends the analysis to nonlinear SAWs in the (111) surface cut. Results are presented for eight different crystals (KCl, NaCl, SrF_2, BaF_2, Si, Ge, Ni, and Cu) over the full range of propagation directions. In addition to the general study of these crystals, detailed studies are presented for the cases of silicon and potassium chloride in the (111) surface cut. In the case of silicon, some of the features described here have been corroborated by previously reported measurements.[3]

For isotropic media, the nonlinearity matrix in the spectral evolution equations has proven to be convenient both to interpret the nature of the spectral interactions[4] and to compute the shock formation distance.[5] In crystalline media, the nonlinearity matrix is especially useful for explaining the different types of waveform distortion that are possible as a result of the reduced symmetry. Unlike in previous theories for isotropic media[6] and for the (001) plane of cubic crystals,[1] the nonlinearity matrix elements in the (111) plane cannot usually be written in real-valued form. In particular, the phase of the nonlinearity matrix elements plays an important role in the resulting waveform distortion. A mathematical transformation is presented which gives reasonably accurate results for the distortion in cases where the phases of the first few nonlinearity matrix elements are similar. From this approach, a graphical table of various types of distortion can be constructed to estimate the nature of the distortion without solving the full system of nonlinear spectral evolution equations for every case.

II. NUMERICAL RESULTS

A. Linear effects

A plane surface wave with wave number k is assumed to propagate in the x direction along the surface of an anisotropic half-space $z \leqslant 0$. The displacement components of the linearized equations of motion can be written in the form [Eq. (22) in Ref. 2]

$$u_j = \sum_{s=1}^{3} C_s \alpha_j^{(s)} \exp[ikl_3^{(s)}] \exp[ik(x-ct)], \tag{1}$$

where $j = x,y,z$, c is the small-signal SAW speed, $l_3^{(s)}$ and $\alpha_j^{(s)}$ are the eigenvalues and eigenvectors, respectively, of the secular equation, and C_s are coefficients which allow the stress-free boundary conditions to be satisfied. The parameters c, $l_3^{(s)}$, $\alpha_j^{(s)}$, and C_s are determined using standard techniques.[7]

The SAW speed c as a function of the propagation direction in selected materials is shown in Fig. 1. These curves were computed using the same data as in Ref. 1. The direction of propagation is measured by the angle θ from the $\langle 11\bar{2} \rangle$ direction. The SAW speeds for each material are scaled by the characteristic speeds $c_{\text{ref}} = (c_{44}/\rho)^{1/2}$, where c_{ij} are the second-order elastic constants in Voigt notation and ρ is the density. The SAW speed is periodic every $\Delta\theta = 60°$ and symmetric about $\theta = 30°$. In most cases (and for all the cases shown here), the speeds group by anisotropy ratio $\eta = 2c_{44}/(c_{11}-c_{12})$, with materials possessing lower anisotropy ratios having higher relative SAW speeds. Materials with $\eta \approx 1$ indicate nearly isotropic media, with a corre-

[a]Current address: Department of Physics, University of Windsor, 401 Sunset Ave, Windsor, ON N9C 4E1, Canada. Electronic mail: kumon@mailaps.org

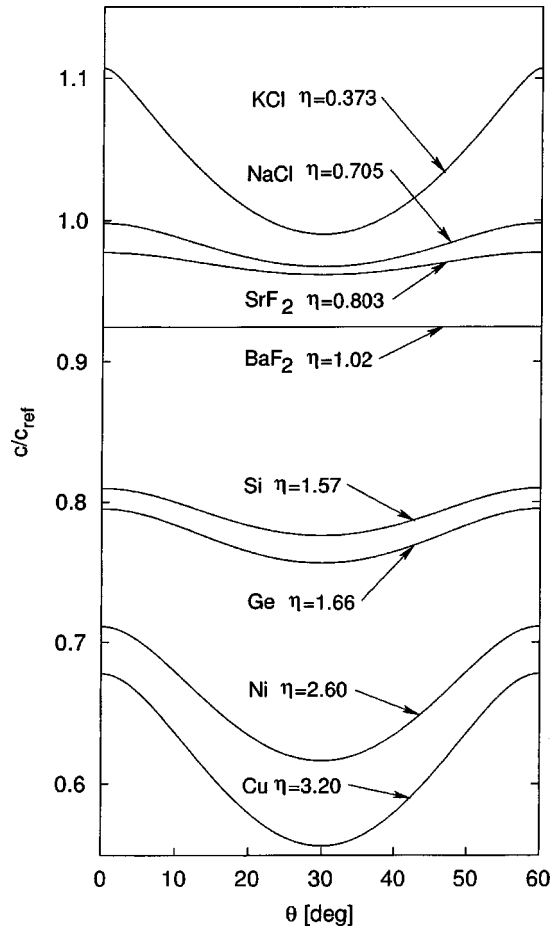


FIG. 1. SAW speed as a function of propagation direction in the (111) plane of selected materials. The SAW speed of each material is measured relative to $c_{\text{ref}}=(c_{44}/\rho)^{1/2}$, and the angle θ gives the direction of the wave vector relative to $\langle 11\bar{2}\rangle$. The wave speeds are periodic every $\Delta\theta=60°$.

spondingly constant SAW speed for all directions. For all materials, the directions $\theta=0°$, $\theta=30°$, and $\theta=60°$ are pure modes, i.e., the wave vector is parallel to the direction of power flow. As $\theta\rightarrow 30°$, the wave speed of the SAW mode tends to approach (but does not equal) the speed of one of the quasitransverse bulk wave modes, and the depth penetration penetration of the SAW mode tends to increase.[7] Unlike in the (001) plane the modes do not converge, and pseudo-surface wave modes do not exist in the direction $\theta=30°$.

While the wave speed has a sixfold periodicity in the (111) plane, other parameters of the linearized SAW equations have only a threefold periodicity.[8] For example, the eigenvalues $l_3^{(s)}$ have only a threefold symmetry. Figure 2 shows the real and imaginary parts of $l_3^{(s)}$ for KCl. (Compare with Fig. 16 in Ref. 7, which shows only $0°\leqslant\theta\leqslant 30°$.) One also finds that $|l_3^{(s)}|$ and $|C_s\alpha_j^{(s)}|$ exhibit a sixfold periodicity, while $\arg[l_3^{(s)}]$ and $\arg[C_s\alpha_j^{(s)}]$ maintain threefold periodicity. Because the nonlinearity matrix is a function of $l_3^{(s)}$ and $C_s\alpha_j^{(s)}$, this result also has implications for the periodicity of nonlinear effects, as discussed below.

B. Nonlinear effects

Because the nonlinear theory used here has been discussed at length elsewhere,[2] only the essential equations are summarized. The coupled nonlinear evolution equations for the surface acoustic waves (without absorption) are[2]

$$\frac{dv_n}{dx}=\frac{n^2\omega}{2\rho c^4}\sum_{l+m=n}\operatorname{sgn}(lm)S_{lm(-n)}v_l v_m, \tag{2}$$

where v_n is the spectral amplitude of the nth harmonic, $\omega=kc$ is the angular frequency, and S_{lm} is the nonlinearity

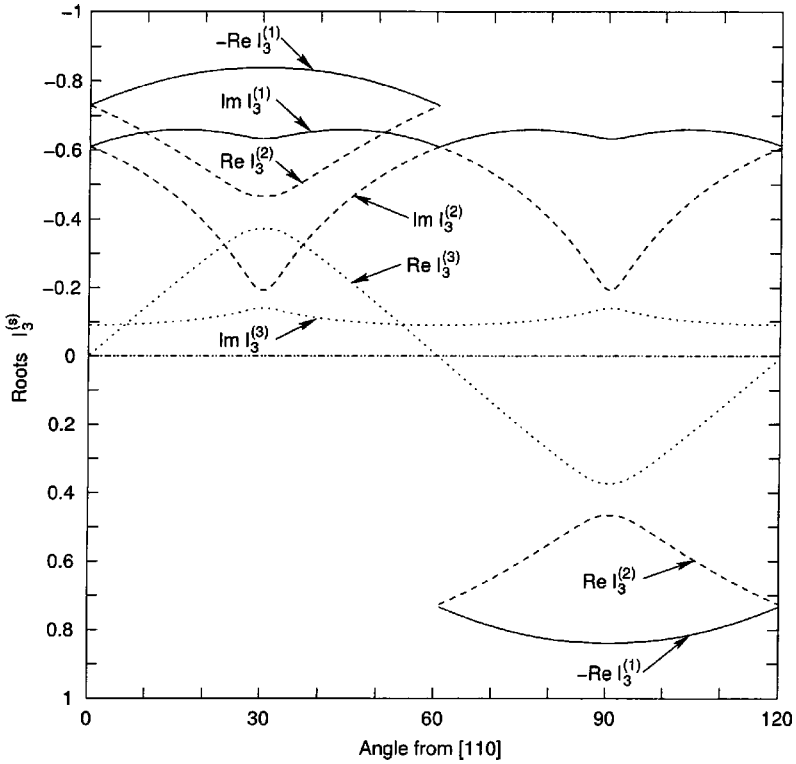


FIG. 2. Dependence of the eigenvalues $l_3^{(s)}$ on the direction of propagation for SAWs in the (111) plane of KCl (solid, long dashed, and short dashed lines for $s=1,2,3$, respectively). The figure is plotted such that it can be directly compared to Fig. 16 in Ref. 7, which only shows $0°\leqslant\theta\leqslant 30°$.

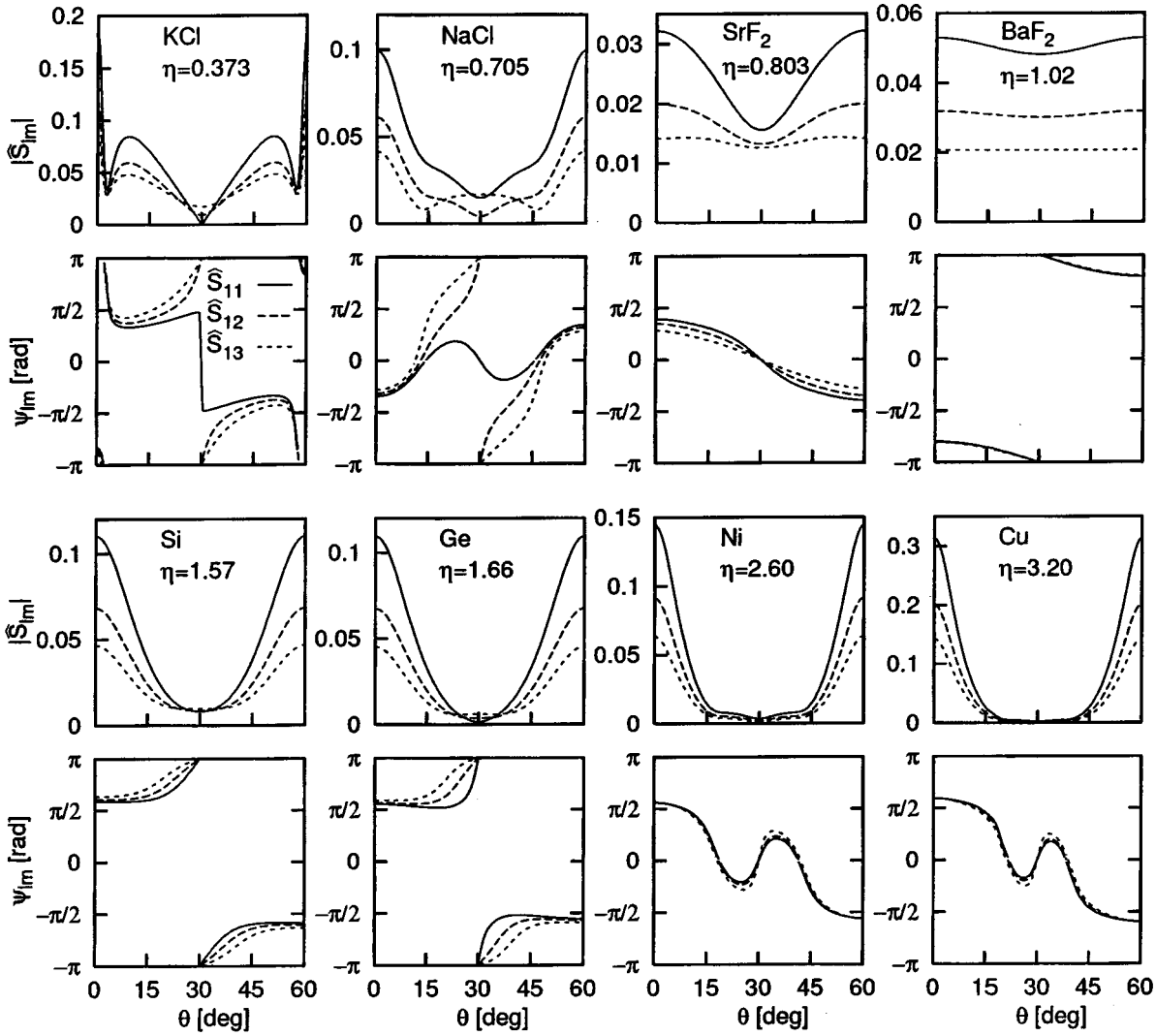


FIG. 3. Dependence of nonlinearity matrix elements on direction of propagation in the (111) plane for selected materials. The solid, long dashed, and short dashed lines correspond to $\hat{S}_{11}$, $\hat{S}_{12}$, and $\hat{S}_{13}$, respectively. The magnitudes $|\hat{S}_{lm}|$ of the matrix elements are plotted in the first and third rows, while the corresponding phases ψ_{lm} are plotted in the second and fourth rows.

matrix. The matrix element S_{lm} describes generation of the nth harmonic due to interaction of the lth and mth harmonics. The velocity waveforms at the surface ($z=0$) are computed from the spectral amplitudes by[1]

$$v_j(x,\tau)=\sum_n v_n(x)|B_j|e^{i\phi_j \operatorname{sgn} n}e^{-in\omega\tau}, \tag{3}$$

where $j=x,y,z$, $B_j=|B_j|\exp(i\phi_j)=C_s\alpha_j^{(s)}$ are determined by solving the linearized equations of motion, and $\tau=t-x/c$ is the retarded time. It is convenient to define a dimensionless nonlinearity matrix[1]

$$\hat{S}_{lm}=-S_{lm}/c_{44}. \tag{4}$$

In all cases, the figures throughout this paper use the nonlinearity matrix defined by Eq. (4).

It is possible that nonlinear coupling between surface wave and bulk wave modes may occur in propagation directions where the wave speeds in these modes are close to one another. The theoretical model does not account for coupling with bulk waves, and this matter was discussed previously in connection with Scholte waves.[9] Calculations have revealed that for the cases described in the present article, effects of mode coupling are negligible in most propagation directions for realistic wave amplitudes. Exceptions may include propagation directions in the region near $\theta=0°$ in Ni and Cu; care should be exercised in applying the aforementioned theory in these cases. A thorough analysis of this matter will be discussed in a future article.

1. General study

Figure 3 displays the nonlinearity matrix elements for KCl, NaCl, SrF_2, BaF_2, Si, Ge, Ni, and Cu. These are the same materials considered in Ref. 1 for the (001) surface cut. Plots for materials with $\eta<1$ or $\eta\approx1$ are given in the top two rows, and for materials with $\eta>1$ in the bottom two rows. Because the matrix elements are usually complex-valued, two plots are given for each material. The top plot in each pair shows the magnitudes $|\hat{S}_{11}|$ (solid), $|\hat{S}_{12}|$ (long dashed), and $|\hat{S}_{13}|$ (short dashed). The inequalities $|\hat{S}_{11}|>|\hat{S}_{12}|>|\hat{S}_{13}|$ hold in most directions within the crystals shown. In these cases, the primary effect of a change in

magnitude over a range of directions is to change the length scale over which the nonlinear distortion occurs. Exceptions occur when one or more of the matrix elements is zero[1] or when one or more of the above inequalities is reversed. The latter case occurs near $\theta=30°$ for KCl, NaCl, Si, and Ge (see Sec. II B 2 for the effect on the waveforms in Si). The matrix elements tend to decrease in magnitude as $\theta\rightarrow 30°$; this trend coincides with the increased depth penetration of the SAW described in Sec. II A. In all cases, the magnitudes $|S_{lm}|$ are periodic every $\Delta\theta=60°$ and symmetric about $\theta=30°$.

The phase ψ_{lm} of the matrix element $\hat{S}_{lm}$ is defined such that $\hat{S}_{lm}=|\hat{S}_{lm}|\exp(i\psi_{lm})$. The bottom plot in each pair shows the phases ψ_{11} (solid), ψ_{12} (long dashed), and ψ_{13} (short dashed). As will be shown in subsequent examples, the primary effect of the phase of the matrix elements is to change the shapes of the various velocity waveform components. Note that the nonlinearity matrix elements are always real-valued ($\psi_{lm}=0$ or $\psi_{lm}=\pm\pi$) at $\theta=30°$ because of the crystalline symmetry in this direction. In contrast to the magnitudes, the phases ψ_{lm} are periodic every $\Delta\theta=120°$ and symmetric about $\theta=60°$.

The nonlinearity matrix elements are functions of the parameters $l_3^{(s)}$ and $C_s\alpha_j^{(s)}$ of the linearized problem.[2] For the crystals shown (all in the $m\bar{3}m$ point group), the magnitudes $|S_{lm}|$, $|l_3^{(s)}|$, and $|C_s\alpha_j^{(s)}|$ have sixfold periodicity in the plane, while the phases $|\psi_{lm}|$, $\arg[l_3^{(s)}]$, and $\arg[C_s\alpha_j^{(s)}]$ have only threefold periodicity. Thus the symmetry properties of the nonlinearity matrix elements are influenced by the symmetry properties of these linear parameters.

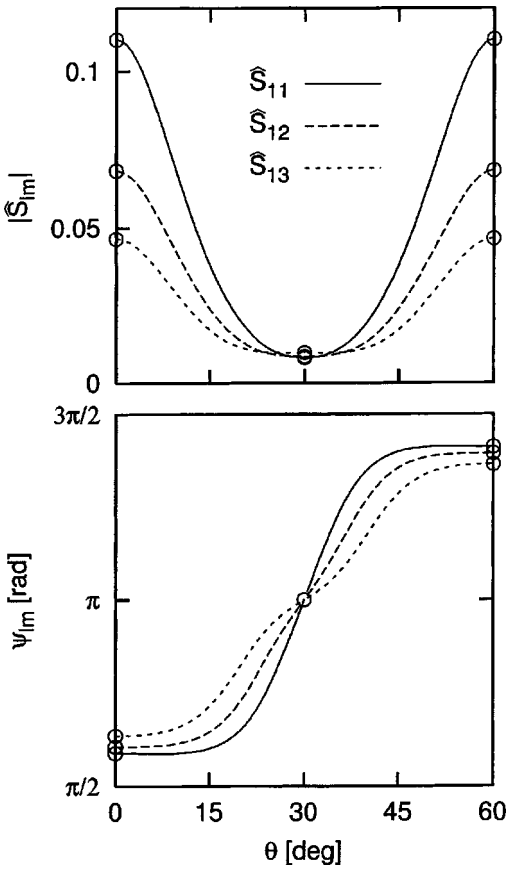


FIG. 4. Nonlinearity matrix elements $\hat{S}_{11}$, $\hat{S}_{12}$, and $\hat{S}_{13}$ for Si in the (111) plane as a function of direction for $0°\leqslant\theta\leqslant 60°$. The circled directions are discussed in detail in the text. Note that the vertical scale on the phase graph is changed to $\pi/2\leqslant\psi_{lm}\leqslant 3\pi/2$ as compared to Fig. 3.

2. Detailed study of silicon

Here we consider nonlinear SAWs in crystalline silicon and show some of the various types of waveform distortion which are possible. Figure 4 (expanded from Fig. 3 with the vertical scale of the phase changed to $\pi/2\leqslant\psi_{lm}\leqslant 3\pi/2$) shows the magnitude and phase of the three nonlinearity matrix elements $\hat{S}_{11}$ (solid), $\hat{S}_{12}$ (long dashed), and $\hat{S}_{13}$ (short dashed). While $\theta=0°$, $\theta=30°$, $\theta=60°$ are pure mode directions, none is a "Rayleigh-type" mode.[7] In the $\theta=0°$ and $\theta=60°$ cases, while the displacement is confined to the sagittal plane, the phase difference between the longitudinal and vertical components is not 90°. Hence the major axis of the initial surface displacement ellipse is not perpendicular to the surface. In the $\theta=30°$ case, the displacement is not confined to the sagittal plane.

Figure 5 displays the velocity waveforms for the directions $\theta=0°$, $\theta=30°$, and $\theta=60°$ marked by small circles in Fig. 4. These waveforms were calculated under the same conditions as described in Ref. 1 and are selected to show the types and diversity of distortion in this cut. The columns from left to right give the dimensionless longitudinal (V_x), transverse (V_y), and vertical (V_z) components of the velocity, respectively. In each direction, the waveforms are normalized such that at $X=0$ we have $|V_x|^2+|V_y|^2+|V_z|^2=1$, and hence the absolute magnitudes between directions should not be compared. The velocity waveforms show results at locations $X=x/x_0=0$ (short dashed), $X=1$ (long dashed), and $X=2$ (solid), where $x_0=\rho c^4/4|S_{11}|\omega v_0$ is a characteristic length scale for nonlinear distortion (approximately equal to the shock formation distance, if shocks form)[1] and v_0 is a characteristic velocity magnitude.

$\theta=0°$: This direction provides the first example of asymmetric distortion. While the velocity waveforms are symmetric about $\omega\tau=0$ at $X=0$, the distortion is asymmetric at $X=1$ and $X=2$, unlike both nonlinear Rayleigh waves[6] and nonlinear SAWs in the (001) plane.[1] For example, a nonlinear Rayleigh wave forms a cusped sawtooth wave with a compression or rarefaction shock in the V_x waveform and a U-shaped wave with a cusped peak in the V_z waveform. However, in this direction the V_x waveform distorts into a U-shaped wave with an asymmetrically cusped peak, while the V_z waveform distorts into a sawtoothlike wave with peaks advancing and troughs receding with respect to the retarded time frame. Measurements of finite-amplitude SAWs at $\theta=0°$ in Si corroborate these results.[3] The source of these differences is the complex-valued nature of the nonlinearity matrix elements. Because a full discussion of the relationship between the phase of the matrix elements and the resulting waveform distortion is given in Sec. III, further analysis of these waveforms is delayed until Sec. IV.

$\theta=30°$: In this direction, the nonlinearity matrix elements are all real-valued. As a result, the distortion is

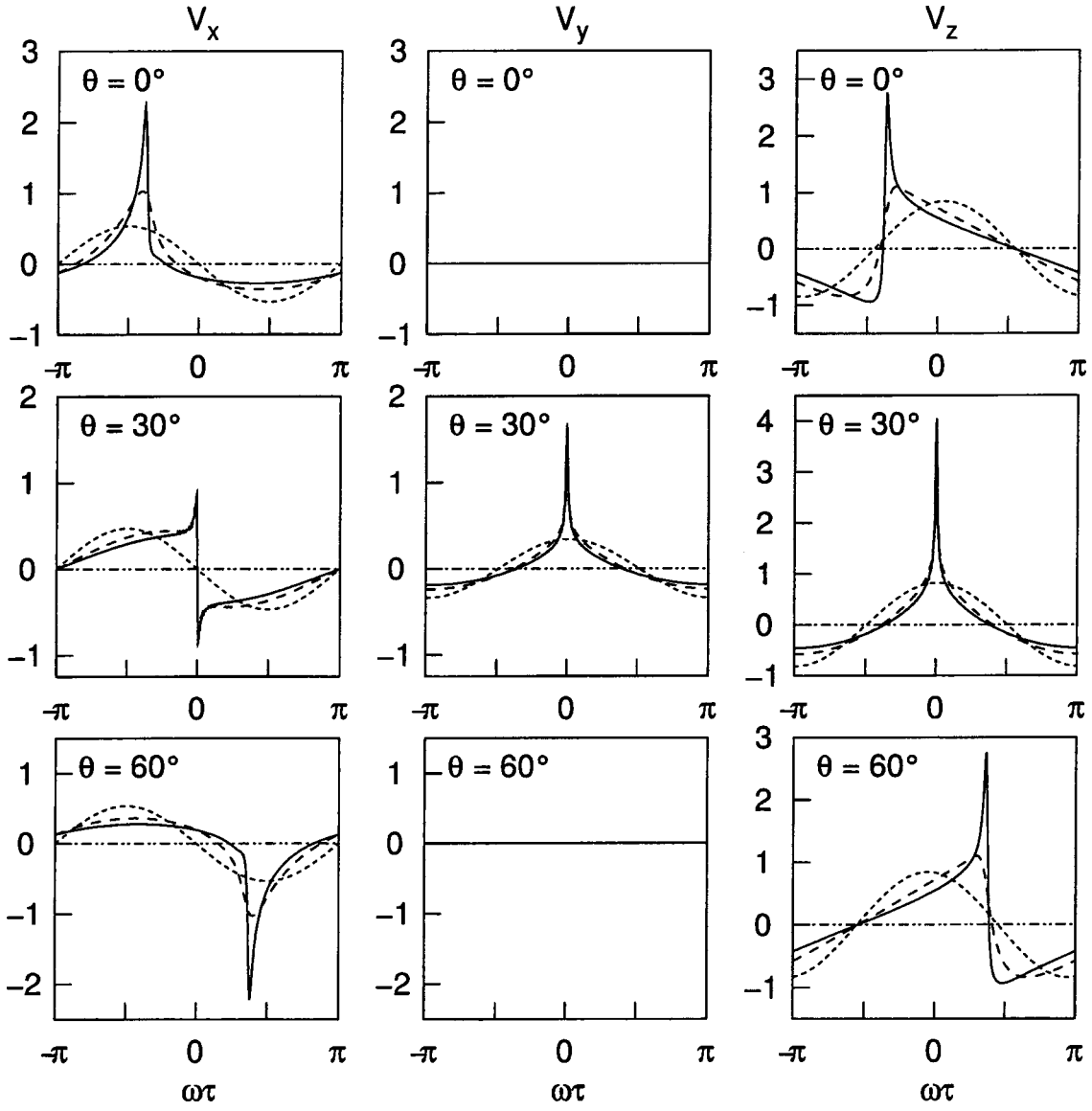


FIG. 5. Velocity waveforms in selected directions for propagation in the (111) plane of Si. The velocity components are normalized such that initial amplitude satisfies $|V_x|^2+|V_y|^2+|V_z|^2=1$ in each propagation direction. The short dashed, long dashed, and solid lines correspond to propagation at distances $X=0$, $X=1$, and $X=2$, respectively.

more similar to nonlinear SAWs in certain directions of the (001) plane.[1] Because $|S_{11}|$ is less than $|S_{12}|$ and $|S_{13}|$, energy is more efficiently transferred from the fundamental to third and higher harmonics than it is to the second harmonic. As a result of the increased energy in the higher harmonics, the velocity waveforms show sharp cusps. In addition, the matrix elements are negative ($\psi_{lm}=\pi$), and so a rarefaction shock forms in the V_x waveform.

$\theta=60°$: While the nonlinearity matrix elements in this direction have the same magnitude as those for $\theta=0°$, their phases have the opposite sign. As a result, the waveforms distort into entirely different shapes. The V_x waveform distorts into an inverted U-shaped wave with an asymmetrically cusped trough, while the V_z waveform distorts into an asymmetric sawtoothlike wave with peaks receding and troughs advancing. Because the nonlinear properties are periodic every $\Delta\theta=120°$, propagation for $\theta=60°$ is the same as for $\theta=180°$. Recently reported measurements[8] have shown that nonlinear propagation is different for $[11\bar{2}]$ and $[\bar{1}\bar{1}2]$ in the (111) plane of crystalline silicon, and the differences in the distortion are consistent with the results shown here.

3. Detailed study of potassium chloride

We consider here nonlinear SAWs in the (111) plane of KCl. The magnitude and phase of $\hat{S}_{11}$ (solid), $\hat{S}_{12}$ (long dashed), and $\hat{S}_{13}$ (short dashed) are shown in Fig. 6 (expanded from Fig. 3 for $0°\leqslant\theta\leqslant30°$ and with the vertical scale of the phase changed to $0\leqslant\psi_{lm}\leqslant3\pi/2$). Like Si, the nonlinearity matrix elements have the largest magnitude at $\theta=0°$. Unlike Si, the phases ψ_{lm} change significantly from $\theta=0°$ to 5°. Here again both the $\theta=0°$ and $\theta=30°$ directions are pure mode directions, but neither is a ''Rayleigh-type'' mode for the same reasons as in Si. One marked difference with Si which occurs even at linear order is that the transverse linear amplitude factors B_y have phases ϕ_y that are

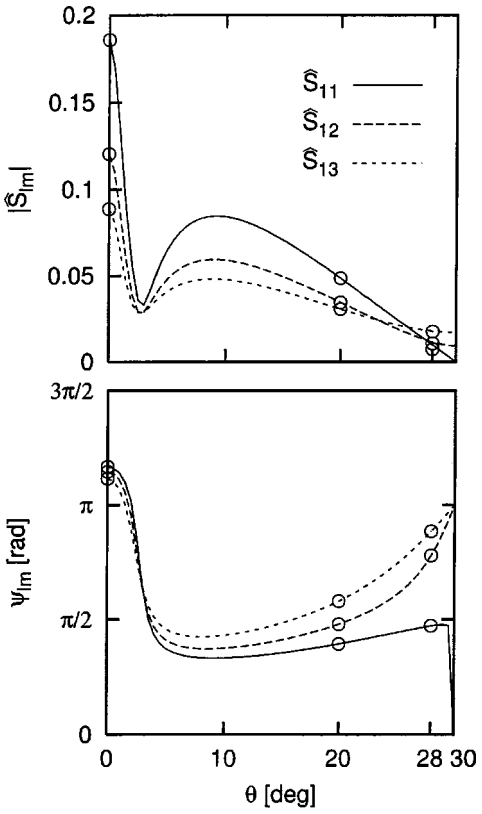


FIG. 6. Nonlinearity matrix elements $\hat{S}_{11}$, $\hat{S}_{12}$, and $\hat{S}_{13}$ for KCl in the (111) plane as a function of direction for $0° \leqslant \theta \leqslant 30°$. The circled directions are discussed in detail in the text. Note that the vertical scale on the phase graph is changed to $0 \leqslant \psi_{lm} \leqslant 3\pi/2$ as compared to Fig. 3.

closer to π than to 0. This change alone results in transverse velocity waveforms that are significantly different.

Figures 7 and 8 display the velocity waveforms and harmonic propagation curves, respectively, for the directions $\theta = 0°$, $\theta = 20°$, and $\theta = 28°$ marked by small circles in Fig. 6. These directions are chosen because they have types of waveform distortion not shown in Fig. 5. Note that the vertical axis for all the waveforms is shifted as compared to Fig. 5 such that $0 \leqslant \omega\tau \leqslant 2\pi$. The harmonic propagation curves show the spectral components V_1 to V_5 as a function of the dimensionless propagation distance. Because the spectral amplitudes are complex-valued, the harmonic propagation curves show both the magnitudes (left column) and phases (right column) of the harmonics. Note that the phases shown are relative to linear theory.

$\theta = 0°$: This direction shows a different type of asymmetric distortion than seen in Si. The V_x waveform forms an asymmetrically cusped sawtoothlike wave with a rarefaction shock and with the negative cusped peak larger in magnitude than the positive cusped peak. The V_z waveform forms a U-shaped wave with an asymmetrically cusped peak. The harmonic magnitude curves in Fig. 8 are typical of nonlinear SAWs in isotropic solids, and the harmonic phase curves show relatively little variation during propagation.

$\theta = 20°$: Observe that the waveforms in this direction distort very differently from those in the $\theta = 0°$ direction. The different shapes result because the dominant nonlinearity matrix elements are clustered near $\pi/2$, instead of near π like at $\theta = 0°$ (see Fig. 6). Moreover, the phases of ψ_{11}, ψ_{12}, and ψ_{13} are more widely spaced at $\theta = 20°$ than at $\theta = 0°$.

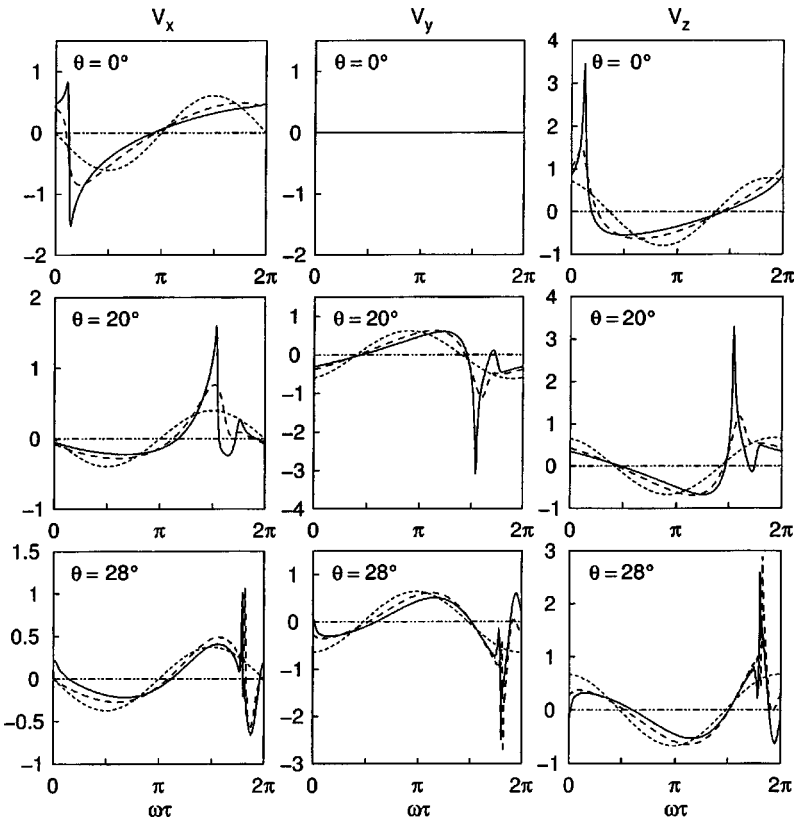


FIG. 7. Velocity waveforms in selected directions of propagation in the (111) plane of KCl. The velocity components are normalized such that initial amplitude satisfies $|V_x|^2 + |V_y|^2 + |V_z|^2 = 1$ in each propagation direction. The short dashed, long dashed, and solid lines correspond to propagation at distances $X = 0$, $X = 1$, and $X = 2$, respectively.

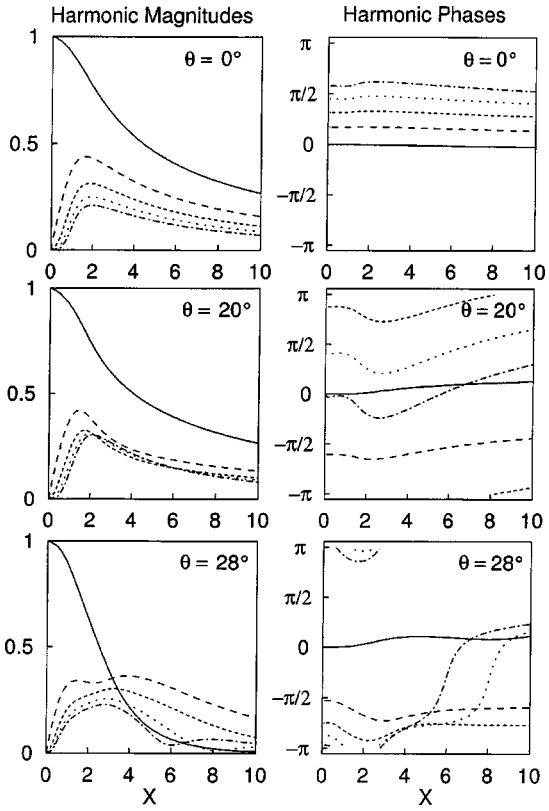


FIG. 8. Harmonic propagation curves for selected directions of propagation in the (111) plane of KCl. The spectral components V_1 (solid), V_2 (long dashed), V_3, (short dashed), V_4 (dotted), and V_5 (dot-dashed) are plotted as a function of distance. The left column shows the spectral amplitudes $|V_n|$, while the right column shows the spectral phases arg V_n relative to linear theory.

This separation causes the oscillations near the shocks and peaks, as discussed further in Sec. IV. While these oscillations are reminiscent of the kind seen in dispersive waves, it should be emphasized that there is no dispersion in this system. In contrast to the $\theta=0°$ direction of KCl, Fig. 8 shows that the magnitudes of the third, fourth, and fifth harmonics become comparable, while the phases are further separated and show larger variation as a function of distance except at the fundamental frequency.

$\theta=28°$: In this direction, $|\hat{S}_{11}|$ is less than $|\hat{S}_{12}|$ and $|\hat{S}_{13}|$, similar to the $\theta=30°$ direction in Si. Energy in the fundamental is thus converted more rapidly to higher harmonics. As shown in Fig. 8, this results in a steeper decline in the magnitude of the fundamental as compared to the previous two directions in KCl, and eventual dominance by the higher harmonics. However, unlike Si, here the phases of the nonlinearity elements are different. Figure 8 shows that the harmonic phases are irregularly spaced and some harmonics change their phase significantly as they propagate. The net result of these complicated interactions is the high-frequency oscillation seen in the waveforms of Fig. 6.

These simulations show that a wide variety of waveform distortion can occur in nonlinear SAWs in the (111) plane as compared to nonlinear Rayleigh waves or nonlinear SAWs in the (001) plane. Features like asymmetric distortion and high-frequency oscillations in the waveforms result from the harmonics having dissimilar phases, which in turn are related to the complex-valued nature of the matrix elements. The next section further investigates the relationship between the nonlinearity matrix elements and waveform distortion.

III. COMPLEX-VALUED NONLINEARITY MATRIX ELEMENTS AND WAVEFORM DISTORTION

Only the concepts of positive and negative nonlinearity are necessary to describe nonlinear SAWs in the (001) plane.[1] Examples of waves with a positive coefficient of nonlinearity β include acoustic waves in fluids, longitudinal bulk waves in most isotropic solids, and SAWs in steel[10] and the direction 26° from ⟨100⟩ in Si.[1] In these waves, the peaks of the longitudinal velocity waveforms advance in a retarded time frame moving at the linear wave speed, while the troughs recede. Examples of waves with $\beta<0$ include SAWs in fused quartz[11] and SAWs propagating in the directions 0° and 35° from ⟨100⟩ in the (001) plane of Si.[1] In these waves, the peaks of the longitudinal velocity waveforms recede in the retarded time while the troughs advance. However, the situation is more complicated for the most general case of a SAW in an anisotropic medium. As shown in Ref. 1, an appropriate coefficient of nonlinearity for SAWs in a crystal is

$$\beta=4c_{44}\hat{S}_{11}/\rho c^2, \tag{5}$$

where the nonlinearity matrix element $\hat{S}_{11}=|\hat{S}_{11}|\exp(i\psi_{11})$ cannot usually be written in real-valued form. The interpretation of Eq. (5) in terms of its effect on waveforms for situations other than $\psi_{11}=0$ (β real and positive) and $\psi_{11}=\pm\pi$ (β real and negative) is not immediately obvious. The purpose of this section is to suggest a way of thinking about this issue.

Ideally, one would like to be able to characterize the type of waveform distortion by computing just a few parameters, thereby avoiding the process of numerically integrating a system of nonlinear differential equations for every material, cut, and direction. As shown in Ref. 1, the nonlinearity matrix elements can serve as such parameters, allowing a reasonable estimate of the type of waveform distortion (or lack thereof) to be determined from plots of the first few elements. The ability to make the same type of estimate is desired here. The specific objective is to investigate in a simplified manner how the phase of the nonlinearity matrix affects the SAW solutions.

Towards this end, the matrix

$$S_{lm}^{\psi}=S_{lm}e^{i\psi\,\mathrm{sgn}\,n} \tag{6}$$

is introduced to represent a nonlinearity matrix constructed by applying a phase increment ψ, independent of l and m, to a given matrix S_{lm}. Given a solution for a material with matrix S_{lm}, it is desired to relate that solution to the one obtained for a material with nonlinearity matrix S_{lm}^{ψ}. It is convenient, although not necessary, to consider the matrix

S_{lm} to be real. The main simplifying assumption is that there exist propagation directions in some materials for which the phase of the nonlinearity matrix is nonzero yet independent of the indices l and m. That Eq. (6) is a reasonable model of the phase dependence in some cases, at least for the first few matrix elements, is supported by Fig. 3. Specifically, one observes that $\psi_{11} \simeq \psi_{12} \simeq \psi_{13}$ for all the materials shown (except SrF_2) at $0° < \theta < 15°$ and $45° < \theta < 60°$. The purpose of introducing $\operatorname{sgn} n$ in Eq. (6) is that S_{lm}^{ψ} must retain all the symmetry properties required of the nonlinearity matrix. In particular, the nonlinearity matrix elements have the symmetry property[2]

$$S_{lm(-n)} = S^*_{(-l)(-m)n}, \tag{7}$$

where $n = l + m$. The $\operatorname{sgn} n$ ensures that Eq. (6) satisfies Eq. (7).

For a nonlinearity matrix S_{lm}^{ψ}, the evolution of nonlinear SAWs is described by Eqs. (2):

$$\frac{dv_n^{\psi}}{dx} = \frac{n^2\omega_0}{2\rho c^4} \sum_{l+m=n} \operatorname{sgn}(lm) S_{lm(-n)}^{\psi} v_l^{\psi} v_m^{\psi}, \tag{8}$$

where the notation v_n^{ψ} designates that these spectral components are the solutions associated with the matrix S_{lm}^{ψ}. Now substitute Eq. (6) into Eq. (8) and multiply both sides by $e^{i\psi \operatorname{sgn} n}$. Let

$$v_n = v_n^{\psi} e^{i\psi \operatorname{sgn} n}, \tag{9}$$

and thus obtain

$$\frac{dv_n}{dx} = \frac{n^2\omega_0}{2\rho c^4} \sum_{l+m=n} \operatorname{sgn}(lm) S_{lm(-n)} v_l v_m. \tag{10}$$

The spectral components v_n in Eq. (10) are recognized as the solutions for a material with nonlinearity matrix S_{lm}. Therefore, the solutions v_n^{ψ} for a material with nonlinearity matrix S_{lm}^{ψ} are related to the solutions v_n for a material with nonlinearity matrix S_{lm} via Eq. (9):

$$v_n^{\psi} = v_n e^{-i\psi \operatorname{sgn} n}. \tag{11}$$

The surface velocity components v_j^{ψ} in the x_j direction for a material with nonlinearity matrix S_{lm}^{ψ} are reconstructed from the spectral components v_n^{ψ} using Eq. (3):

$$v_j^{\psi}(x,\tau) = \sum_n v_n^{\psi}(x) |B_j| e^{i\phi_j \operatorname{sgn} n} e^{-in\omega_0\tau}. \tag{12}$$

For ease of notation, define $\hat{S}_{lm}^{\psi} = -S_{lm}^{\psi}/c_{44}$, following Eq. (4).

Consider an example of the above procedure to relate the phase of the nonlinearity matrix elements to the corresponding type of waveform distortion. Take the well-known waveform distortion for a nonlinear Rayleigh wave with positive nonlinearity coefficient β as a reference case, with spectrum v_n^R, corresponding real-valued nonlinearity matrix elements $\hat{S}_{lm}^R$, and $B_x^R = |B_x^R| e^{-i\pi/2} = -i|B_x^R|$ (this convention for B_j is chosen to be consistent with theory for nonlinear Rayleigh waves in isotropic solids[6]). For simplicity, we consider only the longitudinal velocity waveforms. The longitudinal velocity waveform for the nonlinear Rayleigh wave is given by

$$v_x^R(x,\tau) = \sum_n v_n^R(x)(-i|B_x^R|\operatorname{sgn} n) e^{-in\omega_0\tau}. \tag{13}$$

Now suppose there exists a hypothetical crystal with nonlinearity matrix elements $\hat{S}_{lm}^{\psi} = \hat{S}_{lm}^R e^{i\psi \operatorname{sgn} n}$ and linear amplitude factors $B_x = |B_x| e^{-i\pi/2} = -i|B_x|$. By Eq. (12), the longitudinal velocity waveform at the surface of the crystal is written in terms of the spectral components of the Rayleigh wave as

$$v_x^{\psi}(x,\tau) = \sum_n v_n^{\psi}(x)(-i|B_x|\operatorname{sgn} n) e^{-in\omega_0\tau}, \tag{14}$$

or, from Eq. (11),

$$v_x^{\psi}(x,\tau) = \frac{|B_x|}{|B_x^R|} \sum_n v_n^R(x) e^{-i\psi \operatorname{sgn} n}(-i|B_x^R|\operatorname{sgn} n) e^{-in\omega_0\tau}. \tag{15}$$

Except for the factor of $e^{-i\psi \operatorname{sgn} n}$, the summation is the longitudinal velocity component of the Rayleigh wave. The prefactor $|B_x|/|B_x^R|$ adjusts for possible amplitude differences between the linear solutions of the Rayleigh wave and the SAW in the crystal. Thus, given the linear amplitude factor B_x and the phase ψ of the nonlinearity matrix elements, the waveforms at the surface in the idealized crystal may be computed by changing the phase of the spectral components of the nonlinear Rayleigh wave and reconstructing according to Eq. (15).

The expression in Eq. (15) is only an approximation to the actual waveform. Discrepancies occur because the nonlinearity matrix elements rarely possess identical phase and, even if the phases of the elements are very similar, the magnitudes of the elements may differ. Nevertheless, the overall result can be qualitatively similar, especially in cases where the dominant matrix elements have nearly the same phase.

To gain some intuition about the example transformation $v_n^{\psi} = v_n e^{-i\psi \operatorname{sgn} n}$ given in Eq. (15), the dimensionless waveforms

$$V_x^{\psi}(x,\tau) = \frac{v_x^{\psi}(x,\tau)}{|v_x^{\psi}(0,0)|} = \sum_n v_n^R(x)(-i \operatorname{sgn} n) e^{-i\psi \operatorname{sgn} n} e^{-in\omega_0\tau} \tag{16}$$

are plotted in Fig. 9 for $0 \leqslant \psi \leqslant \pi$. (An analogous figure[12] can be constructed $-\pi \leqslant \psi \leqslant 0$.) The Rayleigh waveforms ($\psi = 0$) are calculated for steel. The third-order elastic constants for the simulations are taken from measurements of "Steel 60 C2H2A" listed in the review by Zarembo and Krasil'nikov,[13] and they are the same as those used in simulations by Zabolotskaya[6] and Shull *et al.*[10] The simulations were performed under conditions identical to the crystal simulations described in Ref. 1. Each plot contains the dimensionless longitudinal velocity waveforms $V_x^{\psi}(X)$ at locations $X = 0$ (short dashed), $X = 1$ (long dashed), and $X = 2$ (solid).

Several observations can be made about this table of graphs. In the limiting cases of $\psi = 0$ and $\psi = \pi$, the waveforms distort positively and negatively, respectively. When

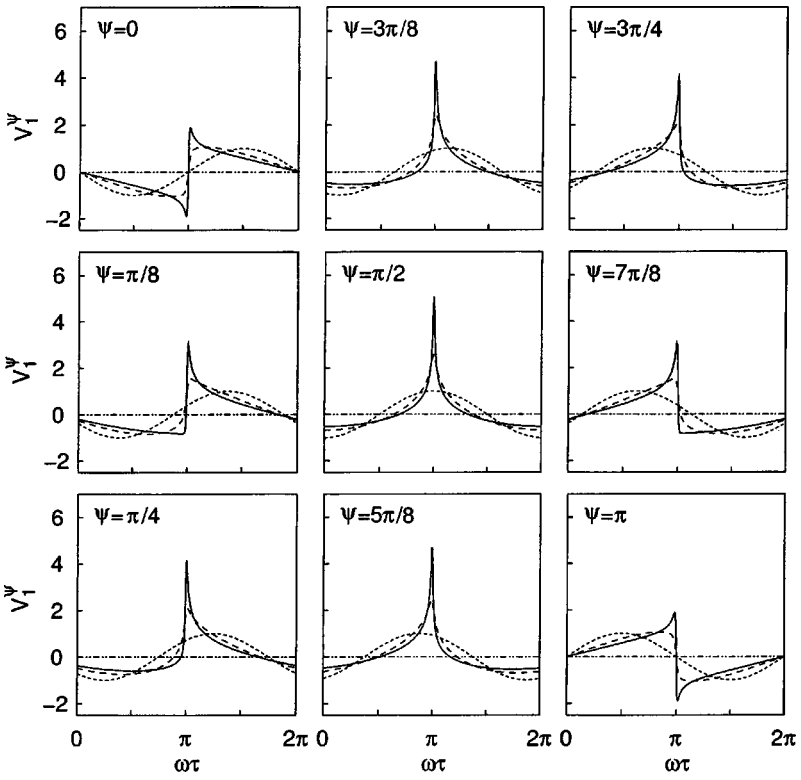


FIG. 9. Transformed waveforms corresponding to the various phase angles $0 \leqslant \psi \leqslant \pi$ of the transformed nonlinearity matrix elements $\hat{S}_{lm}^{\psi} = \hat{S}_{lm} e^{i\psi \operatorname{sgn} n}$, where $n = l + m$. Each graph plots the dimensionless longitudinal velocity waveforms $V_x^{\psi}(x,\tau) = v_x^{\psi}(x,\tau)/|v_x^{\psi}(0,0)|$ for waveforms at distances $X = 0$ (short dashed), $X = 1$ (long dashed), $X = 2$ (solid).

$\psi = \pi/2$, the waveform looks like the vertical velocity component of a negatively distorting Rayleigh wave. This similarity occurs because, from Eq. (11),

$$v_n^{\psi=\pi/2} = v_n e^{-i(\operatorname{sgn} n)\pi/2} = (-i \operatorname{sgn} n) v_n, \tag{17}$$

where the coefficient $-i \operatorname{sgn} n$ is recognized as a Hilbert transform expressed in the frequency domain.[14] The appearance of the Hilbert transform in Eq. (17) is not unexpected because the longitudinal and vertical components of a Rayleigh wave are related in precisely this way. The remaining plots show the expected waveform shapes for intermediate values of ψ and, therefore, other cases of complex-valued nonlinearity matrix elements.

This approximate approach is not limited to longitudinal velocity waveforms. If the phases ϕ_j of the linear amplitude factors B_j are known, then the corresponding approximate waveform distortion in the vertical and transverse directions can also be determined. Let $\psi_{\text{long}} = \psi_{lm}$ be the phase of all the matrix elements. In cases where the phases are not all the same, choose a representative element, typically ψ_{11}. The appropriate values of ψ_{tran} and ψ_{vert} used to characterize the transverse and vertical velocity waveforms are given by[12]

$$\psi_{\text{tran}} = \phi_1 - \phi_2 + \psi_{\text{long}}, \tag{18a}$$

$$\psi_{\text{vert}} = \phi_1 - \phi_3 + \psi_{\text{long}}. \tag{18b}$$

In other words, these are the phases that are appropriate to use with Eq. (12) and Fig. 9 (and its analog for $-\pi \leqslant \psi \leqslant 0$) to determine the approximate shape of the waveform distortion. Calculated phase values using Eqs. (18) for the directions and crystals discussed in this paper are given in Ref. 12.

IV. COMPARISON OF APPROXIMATE AND FULL SOLUTION METHODS

Figure 10 shows comparisons of the transformed solutions based on nonlinear Rayleigh waves in steel and the solutions for the distortion of longitudinal waveforms in several real crystals. The transformed solutions based on Rayleigh waves are constructed via Eq. (12) in three steps. First, the linear amplitude factor B_x and nonlinearity matrix element $\hat{S}_{11}$ are computed for each crystal. Second, the transformation v_n^{ψ} in Eq. (11) is applied to the spectral components v_n^R of the nonlinear Rayleigh waves using ψ_{11}, and the waveforms are translated so that all the sine waves at the source begin in the same place on the horizontal scale. Third, the resulting waveforms are scaled using $|B_x|$ such that the amplitude of the undistorted waveform at $X = 0$ is equal in magnitude to the waveform in the corresponding crystal. The left column gives the transformed Rayleigh wave solutions, while the right column gives the simulations using the full theory (reproduced from Figs. 5 and 7). The rows present comparisons for waveforms propagating in the directions $\theta = 0°$ for Si and $\theta = 0°$, $\theta = 20°$, and $\theta = 28°$ for KCl.

The top two rows show cases for which the nonlinearity elements $\hat{S}_{11}$, $\hat{S}_{12}$, and $\hat{S}_{13}$ have similar, but not the same, phase. For $\theta = 0°$ in Si the characteristic phase was chosen as $\psi_{\text{long}} = \psi_{11} \approx 0.59\pi$ (see Fig. 4). Thus for this case the waveform is expected to be between the shapes given by $\psi = \pi/2$ and $\psi = 5\pi/8$ in Fig. 9. Similarly, the characteristic phase for $\theta = 0°$ in KCl was chosen to be $\psi_{\text{long}} = \psi_{11} \simeq 1.17\pi$ (see Fig. 6). The transformed Rayleigh wave solutions reproduce the general shape of the distortion although not all the details. For example, in the Si waveform the shock

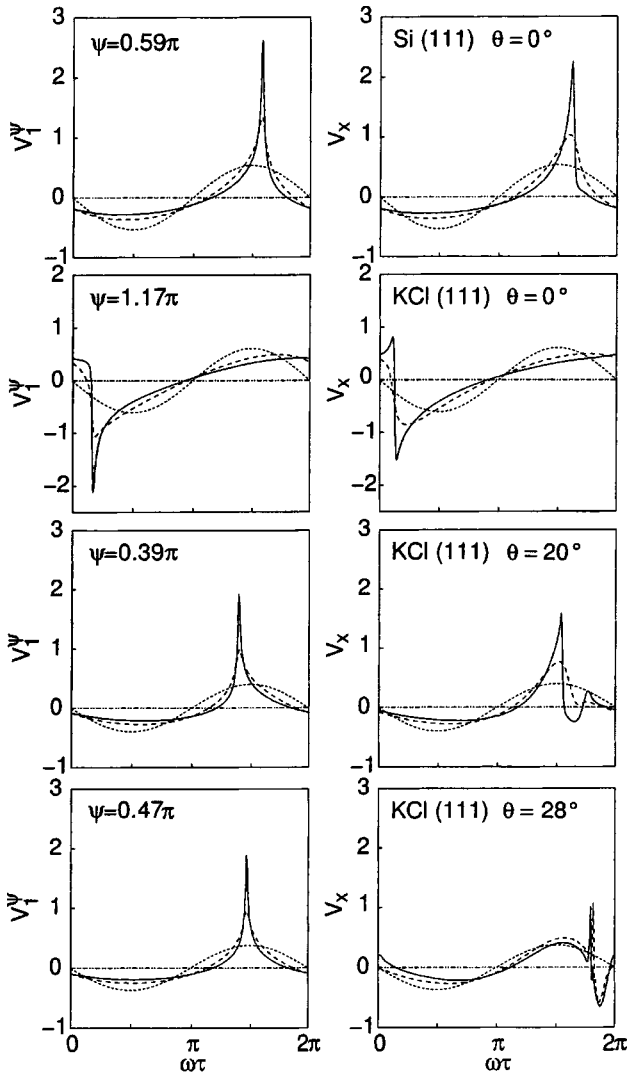


FIG. 10. Comparison of transformed Rayleigh wave solutions (left column) and simulations of nonlinear SAWs in the direction $\theta=0°$ for Si and the directions $\theta=0°$, $\theta=20°$, and $\theta=28°$ for KCl (right column). The top two rows correspond to cases of nonlinearity matrix elements with similar phase, while the bottom two rows show cases where the matrix elements have dissimilar phase.

is steeper than in the transformed waveform, and in the KCl waveform the cusping before the shock does not appear in the transformed waveform. Nevertheless, the similarities between the waveforms are striking. In cases like these, the nonlinearity matrix elements and plots like Fig. 9 can be used to immediately characterize the nature of the waveform distortion.

In contrast, the bottom two rows show several cases where the phases of the nonlinearity matrix elements are less similar. At $\theta=20°$ in KCl, the characteristic phase for the approximate method was chosen to be $\psi_{\text{long}}=\psi_{11}\simeq 0.39\pi$. In this case, the approximately transformed waveforms do not reproduce the extra oscillations that result from the the phase differences introduced between harmonics during the harmonic generation process. However, the occurrence of the oscillations to the right of the cusped peak can be inferred from the relation $\psi_{11}<\psi_{12}<\psi_{13}$. As higher harmonics form, their phase relative to the fundamental is increased. Additional calculations of velocity waveforms in the (111) plane of Ni, where $\psi_{11}>\psi_{12}>\psi_{13}$ (see Fig. 3), correspondingly show that oscillations appear to the left of the peaks and shocks in those waveforms.[12] At $\theta=28°$ in KCl, the characteristic phase for the transformed waveforms was chosen to be $\psi_{\text{long}}=\psi_{11}\simeq 0.47\pi$. However, the approximate method does not reproduce the features seen in the full simulation. This is not unexpected because, as seen in Fig. 6, $|\hat{S}_{11}|<|\hat{S}_{12}|<|\hat{S}_{13}|$, and thus $\hat{S}_{11}$ does not dominate the distortion process. In addition, ψ_{11} differs significantly from ψ_{12} and ψ_{13} in this direction. Under these circumstances, the approximate method is not likely to reproduce all the salient features of the distortion. However, this determination can be made directly by examining a plot of the nonlinearity matrix elements.

Finally, the methods and results described in this paper are not limited to nonlinear SAWs in the (111) surface cut. Additional studies of a variety of cubic crystals in the (110) surface cut have been performed.[12] While the waveform distortion in this cut is similar to that in the (001) plane for crystals in the $m\bar{3}m$ point symmetry group[15] (all the crystals shown in this paper), crystals in the $m\bar{3}$ point group show distortion like in the (111) plane. Examples of crystals in the $m\bar{3}$ point group are the hydrous alums $X\text{Al}(\text{SO}_4)_2\cdot 12\text{H}_2\text{O}$, where $X=$Cs, K, and NH_4. The use of complex-valued nonlinearity matrix elements to characterize waveform distortion is also expected to be applicable to surface waves in materials with other crystalline symmetries besides cubic.

V. SUMMARY

This paper examines the propagation of nonlinear SAWs in the (111) surface cut for a variety of cubic crystals. The SAWs in this plane differ from those in the (001) surface cut in that the nonlinearity matrix elements cannot usually be written in real-valued form. The nonlinearity matrix elements have sixfold symmetry in magnitude but only threefold symmetry in phase. In most directions, initially sinusoidal waveforms distort asymmetrically and, in some cases, the dissimilar phases of the nonlinearity matrix elements result in oscillations forming in the vicinity of the shocks and peaks of the velocity waveforms. Detailed analysis is provided for Si and KCl. A simple mathematical transformation is introduced to provide a graphical interpretation of the phase information contained in the nonlinearity matrix elements and linear amplitude factors. Comparisons are made between waveforms approximated by this method and those generated with the full simulation. The agreement is shown to be best when most of the nonlinearity matrix elements of the crystals have the same or approximately the same phase. By this approach, plots of the nonlinearity matrix elements as a function of direction can be used to characterize the types of harmonic generation and waveform distortion in some directions. The analysis provided here is applicable to other crystals and surface cuts as well, such as in the (110) surface cut of cubic crystals in the $m\bar{3}$ point symmetry group.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. We also gratefully acknowledge discussions with Yu. A. Ilinskii and E. A. Zabolotskaya at The University of Texas at Austin, and P. Hess and A. Lomonosov at the University of Heidelberg, Germany.

[1] R. E. Kumon and M. F. Hamilton, ''Directional dependence of nonlinear surface acoustic waves in the (001) plane of cubic crystals,'' J. Acoust. Soc. Am. **111**, 2060–2069 (2002).

[2] M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, ''Nonlinear surface acoustic waves in crystals,'' J. Acoust. Soc. Am. **105**, 639–651 (1999).

[3] R. E. Kumon, M. F. Hamilton, Yu. A. Il'inskii, E. A. Zabolotskaya, P. Hess, A. Lomonosov, and V. G. Mikhalevich, ''Pulsed nonlinear surface acoustic waves in crystals,'' in *Proceedings of the 16th International Congress on Acoustics and 135th Meeting of the Acoustical Society of America*, edited by P. K. Kuhl and L. A. Crum (Acoustical Society of America, Woodbury, NY, 1998), Vol. 3, pp. 1557–1558.

[4] M. F. Hamilton, Yu. A. Il'insky, and E. A. Zabolotskaya, ''Local and nonlocal nonlinearity in Rayleigh waves,'' J. Acoust. Soc. Am. **97**, 882–890 (1995).

[5] E. Yu. Knight, M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, ''On Rayleigh wave nonlinearity, and analytical approximation of the shock formation distance,'' J. Acoust. Soc. Am. **102**, 2529–2535 (1997).

[6] E. A. Zabolotskaya, ''Nonlinear propagation of plane and circular Rayleigh waves,'' J. Acoust. Soc. Am. **91**, 2569–2575 (1992).

[7] G. W. Farnell, ''Properties of elastic surface waves,'' in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1970), Vol. 6, pp. 109–166.

[8] A. Lomonosov, P. Hess, R. E. Kumon, and M. F. Hamilton, ''Periodicity of linear and nonlinear surface acoustic wave parameters in the (111) plane of cubic crystals,'' J. Acoust. Soc. Am. **110**, 2702(A) (2001).

[9] G. D. Meegan, M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, ''Nonlinear Stoneley and Scholte waves,'' J. Acoust. Soc. Am. **106**, 1712–1723 (1999).

[10] D. J. Shull, M. F. Hamilton, Yu. A. Il'insky, and E. A. Zabolotskaya, ''Harmonic generation in plane and cylindrical nonlinear Rayleigh waves,'' J. Acoust. Soc. Am. **94**, 418–427 (1993).

[11] A. Lomonosov, V. G. Mikhalevich, P. Hess, E. Yu. Knight, M. F. Hamilton, and E. A. Zabolotskaya, ''Laser-generated nonlinear Rayleigh waves with shocks,'' J. Acoust. Soc. Am. **105**, 2093–2096 (1999).

[12] R. E. Kumon, ''Nonlinear surface acoustic waves in cubic crystals,'' Ph.D. dissertation, The University of Texas at Austin, 1999.

[13] L. K. Zarembo and V. A. Krasil'nikov, ''Nonlinear phenomena in the propagation of elastic waves in solids,'' Sov. Phys. Usp. **13**, 778–797 (1971).

[14] R. N. Bracewell, *The Fourier Transform and Its Applications*, 2nd ed. (McGraw–Hill, New York, 1978).

[15] W. Borchardt-Ott, *Crystallography*, 2nd ed. (Springer-Verlag, New York, 1995).

Nonlinear wave interactions in bubble layers

S. Karpov and A. Prosperetti[a)]
Department of Mechanical Engineering, The Johns Hopkins University, Baltimore, Maryland 21218

L. Ostrovsky[b)]
Zel Technologies/NOAA ETL, Boulder, Colorado 80305

(Received 2 May 2001; accepted for publication 30 November 2002)

Due to the large compressibility of gas bubbles, layers of a bubbly liquid surrounded by pure liquid exhibit many resonances that can give rise to a strongly nonlinear behavior even for relatively low-level excitation. In an earlier paper [Druzhinin *et al.*, J. Acoust. Soc. Am. **100**, 3570 (1996)] it was pointed out that, by exciting the bubbly layer in correspondence of two resonant modes, so chosen that the difference frequency also corresponds to a resonant mode, it might be possible to achieve an efficient parametric generation of a low-frequency signal. The earlier work made use of a simplified model for the bubbly liquid that ignored the dissipation and dispersion introduced by the bubbles. Here a more realistic description of the bubble behavior is used to study the nonlinear oscillations of a bubble layer under both single- and dual-frequency excitation. It is found that a difference-frequency power of the order of 1% can be generated with incident pressure amplitudes of the order of 50 kPa or so. It appears that similar phenomena would occur in other systems, such as porous waterlike or rubberlike media. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1539519]

PACS numbers: 43.25.Lj, 43.25.Gf, 43.25.Jh [MFH]

I. INTRODUCTION

The high compressibility of a bubbly mixture causes significant nonlinear effects to arise even at relatively low pressure amplitudes. For this reason, several authors have considered the possibility of using such multiphase systems as active media in parametric arrays (Zabolotskaya and Soluyan, 1973; Kustov *et al.*, 1982; Ostrovsky *et al.*, 1998). The practical realization of this idea has not been very successful due to the use of a mixture of bubble sizes containing resonant bubbles at the incident frequency: the large losses that accompany bubble oscillations near resonance severely diminished the energy available for the parametric effect.

In an earlier paper (Druzhinin *et al.*, 1996) we suggested that the problem could be addressed in a novel way by exploiting the resonances of bubbly liquid layers. It was argued that, by adjusting the bubble size and the operating frequencies so that the desired low-frequency output correspond to layer resonances, it should be possible to operate efficiently while remaining far away from the individual bubble resonance: the result would be an increased efficiency of low-frequency generation and a moderate energy loss. Although promising, it should be noted that the practical realization of this concept requires the generation of bubbles smaller than the resonant radius at the frequencies of interest which, depending on the specifics of the required system, may not be an easy task.

Our earlier work demonstrated the validity of this expectation in principle, but had a preliminary nature in that isothermal gas behavior and a quasi-equilibrium dependence of the bubble radius on the external pressure were assumed. Furthermore, in the presence of the saw-tooth shock wave structure that develops even at moderate pressure amplitudes, the accuracy of the numerical method used in that work was questionable. It is the purpose of this paper to improve on the earlier analysis of the problem, both in the mathematical model and in the numerical treatment. While, unlike the earlier work, the complexity of the model prevents us from obtaining analytic results, the numerical simulations confirm the practical potential of the suggested arrangement.

A schematic representation of the situation studied in this paper is shown in Fig. 1: a one-dimensional layer of liquid containing gas bubbles is located between $x=0$ and $x=L$ and is excited by a plane wave normally incident from the left. As a result of this excitation, a reflected wave at the left of the layer and a transmitted wave at the right are generated.

Some further analysis and preliminary experiments on the low-frequency sound generation in such an arrangement were presented in Ostrovsky *et al.* (1998). The results were however somewhat inconclusive as the layer was resonant only for the difference frequency of the two incident waves, both of which had frequencies close to the individual bubble frequency and, therefore, were strongly dissipated.

II. MATHEMATICAL MODEL

We consider the one-dimensional problem sketched in Fig. 1. The mathematical model of the bubbly liquid consists of the continuity equation

$$\frac{1}{\rho_l c_l^2}\frac{\partial P}{\partial t}+\frac{\partial u}{\partial x}=\frac{\partial \beta}{\partial t}, \tag{1}$$

[a)]Also at Faculty of Applied Physics and Twente Institute of Mechanics, University of Twente, AE 7500 Enschede, The Netherlands, and Burgerscentrum, The Netherlands. Address correspondence to Johns Hopkins address. Electronic mail: prosperetti@jhu.edu
[b)]Also at Institute of Applied Physics, Nizhny Novgorod, Russia.

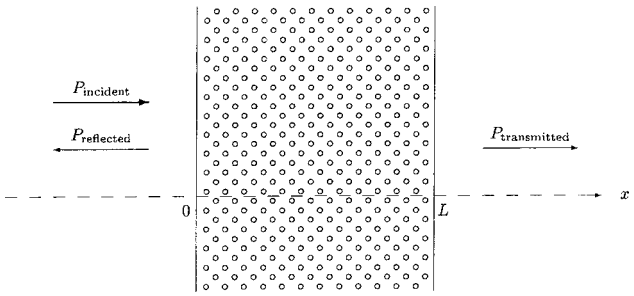


FIG. 1. Schematic representation of the one-dimensional bubble layer excited by a normally incident plane wave from the left.

in which ρ_l and c_l are the density and sound speed of the pure liquid and P and u are the average mixture pressure and velocity, and of the momentum equation

$$\rho_l \frac{\partial u}{\partial t} + \frac{\partial P}{\partial x} = 0. \tag{2}$$

This model is essentially that of Kogarko (1964) and van Wijngaarden (1968) except that, as pointed out by Caflisch *et al.* (1985), the convective term of the material derivative can be omitted due to the assumed smallness of the gas volume fraction β; additional considerations on this point are given in Watanabe and Prosperetti (1994), and further applications of this and similar models can be found e.g., in Zabolotskaya (1977), Kuznetsov *et al.* (1978), Gasenko *et al.* (1979), Nigmatulin (1991), Akhatov *et al.* (1994), Naugolnykh and Ostrovsky (1998), Colonius *et al.* (2000), and many others. Buoyancy effects are neglected in (2) due to the smallness of the acoustic time scale compared with the time evolution of the bubble layer. The volume fraction is given by

$$\beta(x,t) = \tfrac{4}{3}\pi R^3(x,t) n, \tag{3}$$

where $R(x,t)$ is the instantaneous radius of the bubbles contained in a small volume centered around x and n is the bubble number density. In the same assumption $\beta \ll 1$ under which (1) and (2) hold, the bubble number density n can be taken as independent of time; for simplicity, we further assume it to be spatially uniform. The expression (3) can be readily extended to a distribution of bubble sizes by inserting in the right-hand side an integral over the probability distribution of the bubble radii (Zhang and Prosperetti, 1994; Prosperetti, 2001) but, for simplicity, here we assume that all bubbles have the same radius. Generally speaking, if the size distribution is such that all the bubbles have a resonance frequency greater than those of interest, one would not expect very different results (see, e.g., Naugolnykh and Ostrovsky, 1998). On the other hand, if a significant fraction of resonant or near-resonant bubbles were present, dispersion would be very different, dissipation greatly increased, and the phenomena that we discuss strongly and adversely affected.

In spite of its appearance, the previous model retains a strong nonlinearity in the manner in which R is calculated. Again on the basis of the smallness of β, for this purpose we use the Rayleigh–Plesset equation of bubble dynamics (see, e.g., Plesset and Prosperetti, 1977; Prosperetti, 1991; Feng and Leal, 1997):

$$R\frac{\partial^2 R}{\partial t^2} + \frac{3}{2}\left(\frac{\partial R}{\partial t}\right)^2 = \frac{1}{\rho_l}\left(p - P - \frac{2\sigma}{R} - \frac{4\mu}{R}\frac{\partial R}{\partial t}\right). \tag{4}$$

Here p is the bubble internal pressure (approximated by the gas pressure, the small vapor contribution being neglected), σ is the surface tension coefficient, and μ is the liquid viscosity. For an isolated bubble, the ambient pressure P appearing in (4) is to be identified with the pressure at the location of the bubble if the bubble were absent. In a dilute mixture the bubbles are subject to the averaged field and P should be taken as the average pressure appearing in the momentum equation (2) (see, e.g., Caflisch *et al.*, 1985; Zhang and Prosperetti, 1994). As before, in (4) we omit the convective term of the material derivatives of R.

In writing (3) and (4) we have implicitly taken the bubble to be spherical, ignoring the distortion due to gravity, flow, and bubble–bubble interaction. This we do for simplicity and with little loss of accuracy since shape modes couple inefficiently to pressure perturbations when the spherical shape is stable. We have also neglected the corrections due to liquid compressibility (see, e.g., Prosperetti and Lezzi, 1986) which we had included in the early calculations for this paper but found to have a very small influence on the results [see Prosperetti (1984) for a comparison of the various damping mechanisms].

In order to close the system a relationship between the gas pressure p and bubble radius R is needed. This point has been treated at length in earlier papers (Prosperetti, 1991; Watanabe and Prosperetti, 1994). Suffice it to say that we approximate the gas pressure inside each bubble as spatially uniform, which leads to

$$\frac{\partial p}{\partial t} = \frac{3}{R}\left[(\gamma - 1)k\frac{\partial T}{\partial r}\bigg|_R - \gamma p\frac{\partial R}{\partial t}\right], \tag{5}$$

where T is the local gas temperature to be found from

$$\frac{\gamma}{\gamma - 1}\frac{p}{T}\left(\frac{\partial T}{\partial t} + v\frac{\partial T}{\partial r}\right) = \dot{p} + \frac{1}{r^2}\frac{\partial}{\partial r}\left(kr^2\frac{\partial T}{\partial r}\right), \tag{6}$$

with

$$v = \frac{1}{\gamma p}\left[(\gamma - 1)k\frac{\partial T}{\partial r} - \frac{1}{3}r\dot{p}\right]. \tag{7}$$

In these equations γ and $k = k(T)$ are the ratio of specific heats and thermal conductivity of the gas and r is the radial coordinate measured from the center of the bubble. As shown in Kamath *et al.* (1993), at the surface of the bubble, a suitable boundary condition for (6) is

$$T(R,t) = T_\infty, \tag{8}$$

where T_∞ is the undisturbed liquid temperature. It should be noted that, inside the bubble centered at x, the temperature field T depends on r as well as t and, hence, in principle, the set of equations (5)–(7) must be solved at all spatial locations in the layer.

In Druzhinin *et al.* (1996) we considered a much simpler model of the bubble behavior that was obtained from (4) by assuming an instantaneous equilibrium between internal and external pressures and ignoring viscosity and surface tension so that $p=P$; in addition, the bubble internal pressure was related to the radius by a polytropic assumption so that Eqs. (1) and (2) were closed by the simple relation

$$p_0\left(\frac{R_0}{R}\right)^{3\kappa_p}=P, \tag{9}$$

with κ_p a polytropic index. While the neglect of viscosity and surface tension is not very limiting for bubble sizes above a few tens of micrometers, the neglect of inertia on the left-hand side of (4) restricts the validity of this quasi-equilibrium model to frequencies much lower than the bubble resonance frequency. In this limit, in most parameter regions of interest, the bubbles behave isothermally so that Eq. (9) is justified with $\kappa_p=1$. Outside this very restricted domain of validity, however, the model of our earlier paper cannot be expected to be accurate.

A model intermediate between the quasi-equilibrium and complete models can be formulated by using, in the Rayleigh–Plesset equation (4), the polytropic relation (9) in place of (5) for the bubble internal pressure. This model, termed *polytropic* in the following, accounts for the inertia of the bubble radial motion, but not for the strongly dissipative thermal effects.

A. Boundary conditions

As the layer compresses and expands under the action of an incident pressure wave, the planes defining its boundaries will move normal to themselves. If the gas volume fraction is small, by the same argument that enables us to neglect the convective term in the material derivatives, we may disregard this effect and approximate the layer boundaries as fixed in space. As shown in Druzhinin *et al.* (1996), this approximation permits a great simplification of the problem.

Indeed, since the medium outside the bubble layer can be regarded as linear and is nondispersive, one may assume that the incident, reflected, and transmitted waves have the form

$$\begin{aligned} &P_{\rm inc}=P_i(x-c_l t), \quad P_{\rm trans}=P_t(x-c_l t),\\ &P_{\rm ref}=P_r(x+c_l t). \end{aligned} \tag{10}$$

At the layer boundaries the pressure should be continuous so that if, as before, P denotes the pressure in the layer,

$$P_i(0,t)+P_r(0,t)=P(0,t), \quad P(L,t)=P_t(L,t). \tag{11}$$

The velocity or, equivalently, the pressure gradients should also be continuous which, at $x=0$, gives

$$\frac{\partial P_i}{\partial x}+\frac{\partial P_r}{\partial x}=\frac{\partial P}{\partial x}. \tag{12}$$

By virtue of the particular form (10) of the functional dependence on x and t of P_i and P_r, the spatial derivatives are readily related to time derivatives and, upon taking the time derivative of (11), we can eliminate $\partial P_r/\partial t$. The result is the condition

$$\frac{\partial P}{\partial t}-c_l\frac{\partial P}{\partial x}=2\frac{\partial P_i}{\partial t} \quad \text{at } x=0. \tag{13}$$

Proceeding similarly at the right boundary we find

$$\frac{\partial P}{\partial t}+c_l\frac{\partial P}{\partial x}=0 \quad \text{at } x=L. \tag{14}$$

Because of this argument, the problem is reduced to solving the equations of Sec. II subject to the conditions (13) and (14) at $x=0$ and $x=L$. After the pressure P inside the bubble layer is determined, the reflected and transmitted waves are found from (11).

III. LINEAR RESULTS

For a better understanding of the material that follows, it is useful to summarize here some results of the linear analysis of the previous model.

As shown in Commander and Prosperetti (1989), for linear pressure perturbations the model of the previous section gives the following dispersion relation for monochromatic pressure waves with a time dependence proportional to $\exp(i\omega t)$ and wave number κ in a bubbly liquid:

$$\frac{\kappa^2}{\omega^2}=\frac{1}{c_l^2}+\frac{3\beta_0/R_0^2}{\omega_0^2-\omega^2+2ib\omega}. \tag{15}$$

A relation of this type has been derived by many authors starting from several similar models (see, e.g., Carstensen and Foldy, 1947; Clay and Medwin, 1977; Waterman and Truell, 1961; Twersky, 1962; Omta, 1987; d'Agostino and Brennen, 1988; Nigmatulin, 1991; Medwin and Clay, 1997). Here

$$\beta_0=\tfrac{4}{3}\pi R_0^3 n \tag{16}$$

is the gas volume fraction at equilibrium, $\omega_0/2\pi$ is the effective undamped resonance frequency of the bubbles, and b is the effective damping parameter; these quantities depend on the driving frequency ω and are given by

$$\omega_0^2=\frac{p_0}{\rho_l R_0^2}\left(3\kappa_l-\frac{2\sigma}{R_0 p_0}\right), \quad \kappa_l=\frac{1}{3}\,\mathrm{Re}\,\Phi, \tag{17}$$

$$b=\frac{2\mu_l}{\rho_l R_0^2}+\frac{p_0}{2\rho_l R_0^2\omega}\,\mathrm{Im}\,\Phi, \tag{18}$$

where κ_l is the linear-theory value of the polytropic index which is calculated from the complex function Φ defined by (Prosperetti, 1991)

$$\Phi=\frac{3\gamma}{1-3(\gamma-1)iz[(i/z)^{1/2}\coth(i/z)^{1/2}-1]}, \tag{19}$$

with $z=D/\omega R_0^2$, in which D is the gas thermal diffusivity. The ratio $c_m=\omega/\kappa$ is the phase velocity of the wave which (15) shows to be complex: the imaginary part describes the attenuation of the wave in the bubbly mixture due to the energy losses in the bubbles. In a water–air system viscous losses, described by the first term on the right-hand side of (18), are much smaller than the thermal ones except for bubble radii in the micrometer range. If liquid compressibility effects were retained in (4), an acoustic loss contribution

$b_{ac}=\omega^2 R_0/(2c_l)$ would be added to the right-hand side of (18); this term is also small for ω below the bubble resonance frequency ω_0.

For the quasi-equilibrium model of Eq. (9) the relation corresponding to (15) is

$$\frac{\kappa^2}{\omega^2}=\frac{1}{c_l^2}+\frac{\beta_0\rho_l}{\kappa_p P_\infty}, \tag{20}$$

where P_∞ is the undisturbed pressure in the liquid. This relation demonstrates the extreme sensitivity of the effective speed of sound κ/ω to the gas concentration; for example, in water $\rho_l c_l^2/P_\infty \simeq 2.25\times 10^4$ and, with $\kappa_p=1$ and $\beta=10^{-3}$, we have from (20) $c_l/c_m \simeq 5$.

For a monochromatic wave with unit amplitude normally incident on a bubble layer of thickness L, it is easy to show that the amplitudes of the transmitted and reflected waves, $A_{\rm tr}$ and $A_{\rm ref}$, are given by standard acoustic relations (see, e.g., Pierce, 1989, Secs. 3–7; Commander and Prosperetti, 1989)

$$A_{\rm tr}=\frac{\exp(i\omega L/c_l)}{\cos(\kappa L)+\frac{1}{2}i[\omega/(\kappa c_l)+\kappa c_l/\omega]\sin(\kappa L)}, \tag{21}$$

$$A_{\rm ref}=\frac{\frac{1}{2}i[\omega/(\kappa c_l)-\kappa c_l/\omega]\sin(\kappa L)}{\cos(\kappa L)+\frac{1}{2}i[\omega/(\kappa c_l)+\kappa c_l/\omega]\sin(\kappa L)}. \tag{22}$$

The corresponding results for the quasi-equilibrium model are conveniently written in the following form:

$$A_{\rm tr}=\frac{4A}{(A+1)^2-(A-1)^2\exp(-2iA\omega L/c_l)}, \tag{23}$$

$$A_{\rm ref}=\frac{(A^2-1)[\exp(-2iA\omega L/c_l)-1]}{(A+1)^2-(A-1)^2\exp(-2iA\omega L/c_l)}, \tag{24}$$

where the nondimensional parameter A is defined by

$$A=\frac{c_l}{c_m}=\sqrt{1+\frac{\rho_l\beta_0 c_l^2}{P_\infty\kappa_p}} \tag{25}$$

with the second equality following from (20). Equations (23) and (24) give a simple estimate of the resonance frequencies of the incident wave:

$$\omega_n=\frac{n\pi c_l}{AL}, \quad n=1,2,\ldots. \tag{26}$$

This estimate can be refined by calculating numerically the maxima and minima of the moduli of $A_{\rm tr}$ and $A_{\rm ref}$ given by (21) and (22).

IV. NUMERICAL METHOD

The problem to be solved can be decomposed in two components, the integration of the continuity and momentum equations in the bubbly liquid, and the calculation of the temperature inside the bubbles. While these two components are coupled, their nature is very different and so must be their numerical treatment.

In view of the strong nonlinear effects in the bubble layer, steep waveforms develop in the system. To avoid the well-known numerical oscillations that can arise in these conditions, we use the total variation diminishing method of Harten (1983), the implementation of which is now briefly described. It is convenient to use dimensionless variables defined according to

$$\begin{aligned} &x'=\frac{x}{L}, \quad t'=\frac{c_l t}{L}, \quad u'=\frac{\rho_l c_l}{P_\infty}u,\\ &P'=\frac{P}{P_\infty}-1, \quad \beta'=\frac{\beta}{\beta_0}, \end{aligned} \tag{27}$$

but we drop the primes for convenience.

The continuity and momentum equations (1) and (2) may be compactly written as

$$\frac{\partial \mathbf{w}}{\partial t}+\frac{\partial \mathbf{F}}{\partial x}=\frac{\rho_l\beta_0 c_l^2}{P_\infty}\mathbf{b}, \tag{28}$$

where the vectors $\mathbf{w}$, $\mathbf{F}$, and $\mathbf{b}$ are given by

$$\mathbf{w}=\begin{vmatrix} P \\ u \end{vmatrix}, \quad \mathbf{F}=\begin{vmatrix} u \\ P \end{vmatrix}, \quad \mathbf{b}=\begin{vmatrix} \partial\beta/\partial t \\ 0 \end{vmatrix}. \tag{29}$$

This system is discretized explicitly in time and in space as

$$\frac{\mathbf{w}_i^{n+1}-\mathbf{w}_i^n}{\Delta t}+\frac{\hat{\mathbf{F}}_{i+1/2}^n-\hat{\mathbf{F}}_{i-1/2}^n}{\Delta x}=\frac{\rho_l\beta_0 c_l^2}{P_\infty}\mathbf{b}_i^{n+1/2}, \tag{30}$$

where superscripts indicate time levels and subscripts spatial nodes. The modified fluxes $\hat{\mathbf{F}}$ are given by the following expression:

$$\begin{aligned} \hat{\mathbf{F}}_{i+1/2}=&\frac{1}{2}(\mathbf{F}_{i+1}+\mathbf{F}_i)\\ &+\frac{s}{2}\sum_{l=1}^{2}[g_i^{(l)}+g_{i+1}^{(l)}-\alpha_{i+1/2}^{(l)}Q(sa^{(l)}+\eta_{i+1/2}^{(l)})]\mathbf{R}^{(l)}, \end{aligned} \tag{31}$$

where $s=c_l\Delta t/\Delta x$ is the Courant number and

$$\mathbf{R}^{(1)}=\begin{vmatrix} 1 \\ 1 \end{vmatrix}, \quad \mathbf{R}^{(2)}=\begin{vmatrix} 1 \\ -1 \end{vmatrix}, \quad a^{(1)}=1, \quad a^{(2)}=-1, \tag{32}$$

$$\begin{aligned} \alpha_{i+1/2}^{(1)}&=\tfrac{1}{2}(P_{i+1}-P_i+u_{i+1}-u_i),\\ \alpha_{i+1/2}^{(2)}&=\tfrac{1}{2}(P_{i+1}-P_i-u_{i+1}+u_i), \end{aligned} \tag{33}$$

$$\eta_{i+1/2}^{(l)}=\begin{cases} (g_{i+1}^{(l)}-g_i^{(l)})/\alpha_{i+1/2}^{(l)} & \text{if } \alpha_{i+1/2}^{(l)}\neq 0,\\ 0 \quad \text{if } \alpha_{i+1/2}^{(l)}=0. \end{cases} \tag{34}$$

The function Q is defined by

$$Q(y)=\begin{cases} y^2/(4\epsilon)+\epsilon, & \text{for } |y|<2\epsilon,\\ |y|, & \text{for } |y|\geqslant 2\epsilon, \end{cases} \tag{35}$$

with $\epsilon=0.1$, and plays the role of an artificial viscosity. Furthermore,

$$g_i^{(l)}=\begin{cases} \min(|[Q(s)-s^2]\alpha_{i+1/2}^{(l)}|,|[Q(s)-s^2]\alpha_{i-1/2}^{(l)}|)\\ \qquad \text{if } \alpha_{i+1/2}^{(l)}\alpha_{i-1/2}^{(l)}\geqslant 0,\\ 0 \quad \alpha_{i+1/2}^{(l)}\alpha_{i-1/2}^{(l)}<0. \end{cases} \tag{36}$$

The terms in the summation in (31) are a correction to the components of the flux $\mathbf{F}$ along the characteristic directions,

which are introduced to account for the discretization error and guarantee second-order accuracy in space.

The time integration is a variant of a predictor-corrector method abbreviated so as to result in a faster execution without significant loss of accuracy. Briefly, the procedure is as follows. Suppose that everything is known at time level t^n. At all interior nodes we generate a preliminary estimate of the pressure $\tilde{P}^{n+1}$ at time level $t^{n+1}=t^n+\Delta t$ from (30) in which $(\partial_t\beta)^n$ is used in the vector **b** on the right-hand side (here and in the following, tildes denote provisional estimated values at time t^{n+1}.) With this updated pressure we calculate new values of $(\partial_t^2\tilde{R})^{n+1}$ and $(\partial_t\tilde{R})^{n+1}$ and use an approximation to the trapezoidal rule in the form

$$R^{n+1}=R^n+\tfrac{1}{2}\Delta t[(\partial_t R)^n+(\partial_t\tilde{R})^{n+1}], \tag{37}$$

and similarly for all the other variables. In executing this step, in principle it would be necessary to update the time derivative of the bubble pressure (5) as well, which would however significantly slow down the calculation. We found that simply using the value $\partial_t p|^n$ does not lead to a significant loss of accuracy.

After a variable transformation that fixes the bubble boundary, the energy equation in the gas (6) is turned into a set of ordinary differential equations in time by the Galerkin spectral method described in Kamath and Prosperetti (1989). The temperature is expanded over a set of even Chebyshev polynomials; the number of polynomials used in the expansion varies in time according to the procedure given in Kamath and Prosperetti (1989).

To generate the numerical results that follow we have typically used between 800 and 3200 spatial nodes per wavelength depending on the amplitude and the volume fraction: stronger incident waves lead to shock formation the resolution of which requires more nodes. The appropriate number of nodes was chosen by successively refining the grid until the results stabilized. For each level of discretization the time step Δt was chosen such that the Courant number was less than 0.2.

It is clear from the preceding description that the problem is solved as an initial-value problem. The time necessary to reach a steady state depends on the driving amplitude and the value of the parameter A. For weak excitation one typically needs about $20L/\lambda_m$ cycles, where λ_m is the wavelength in the bubbly mixture. For larger amplitudes dissipation is stronger and the numerical constant of 20 can be considerably reduced (Druzhinin *et al.*, 1996).

V. RESULTS: SINGLE-FREQUENCY EXCITATION

In all the examples that follow we use the physical properties of an air–water system at standard conditions. Specifically, we take $P_\infty=100$ kPa, $\rho_l=10^3$ kg/m^3, $c_l=1.5\times10^3$ m/s, and $\mu_l=10^{-3}$ N s/m^2, $\gamma=1.4$, $\sigma=0.07$ N/m; the air thermal conductivity is calculated from $k(T)=A_K T+B_K$ with $A_K=5.528\times10^{-5}$ J/(m s K^2) and $B_K=1.165\times10^{-2}$ J/(m s K), which provides a good fit to the data in the range 200 K$<T<$3000 K. The width of the bubble layer is taken to be $L=0.1$ m and the coupling parameter A defined in Eq. (25), evaluated for $\kappa_p=1$, is taken to be $\sqrt{31}\simeq5.57$, which corresponds to a bubble volume fraction $\beta_0=0.133\%$. Furthermore, we estimate the bubbly layer lowest mode ω_1 from (26) with $n=1$ and $\kappa_p=1$.

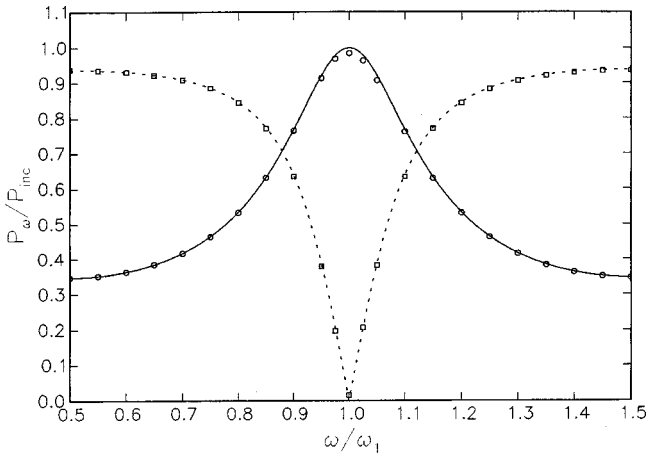


FIG. 2. Pressure amplitudes of the component at the incident wave frequency ω for the transmitted (circles) and reflected (squares) waves as functions of the incident frequency ω normalized by the first linear eigenfrequency of the layer according to the quasi-equilibrium model, $\omega_1/2\pi=1.347$ kHz; the solid and dotted lines are the linear results given in Eqs. (23) and (24). The incident wave amplitude $P_{\rm inc}/P_\infty=0.05$ is smaller than the threshold value for shock formation given by (38). The gas volume fraction is $\beta_0=0.133\%$, the ambient pressure $P_\infty=100$ kPa, the liquid density $\rho_l=10^3$ kg/m^3, the liquid speed of sound $c_l=1.5\times10^3$ m/s, the liquid viscosity $\mu_l=10^{-3}$ N s/m^2, the gas adiabatic index $\gamma=1.4$, the surface tension coefficient $\sigma=0.07$ N/m, and the width of bubble layer $L=0.1$ m.

Before turning to the results of the complete model, it is useful to study those given by the quasi-equilibrium model (9) for which the earlier results of Druzhinin *et al.* (1996), as well as the analytical ones of Sec. III, are available; clearly, the predictions of the quasi-equilibrium model are independent of the bubble radius.

A. Quasi-equilibrium model

Druzhinin *et al.* (1996) give the following criterion for the threshold amplitude of an incident monochromatic wave that leads to shock formation in the layer:

$$\frac{P_{\rm inc}^{\rm th}}{P_\infty}=\frac{4J_2(2)}{\pi J_1(1)}\frac{1}{n\sqrt{A^2-1}}, \tag{38}$$

where $J_{1,2}$ are Bessel functions, and it is assumed that the incident frequency corresponds to the nth resonance frequency of the layer. For $A=\sqrt{31}$, $n=1$, and isothermal oscillations, this relation gives a value of 0.18. Thus, in order to start with the linear regime, we consider first a case with an incident wave amplitude $P_{\rm inc}/P_\infty=0.05$. Figure 2 shows the transmitted (circles) and reflected (squares) amplitudes as functions of the ratio ω/ω_1 of the incident frequency to the lowest layer eigenfrequency which, here, is $\omega_1/2\pi=1.347$ kHz as given by (26). The amplitudes shown are for the components at the same frequency as the incident wave, which are found by taking the Fourier transforms of the total transmitted and reflected waves; the relative power of the higher-frequency components is smaller than 0.0004. In the figure the lines are the analytical results (23) and (24). The agreement is excellent, which suggests that the numerical

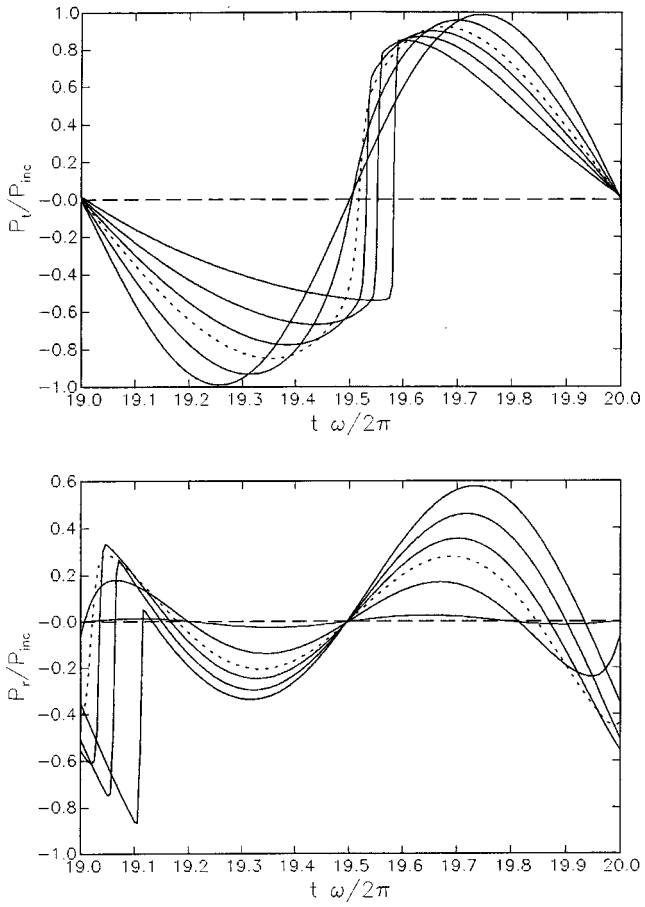


FIG. 3. Steady-state shape of transmitted (upper panel) and reflected waves for different amplitudes of the incident wave, $P_{\rm inc}/P_\infty=0.01$, 0.1, 0.2 (dotted), 0.3, 0.5, 0.9 according to the quasi-equilibrium model. The frequency of the incident wave is equal to the first linear eigenfrequency of the layer $\omega=\omega_1$, with $\omega_1/2\pi=1.347$ kHz. The gas volume fraction is $\beta_0=0.133\%$, the ambient pressure $P_\infty=100$ kPa, the liquid density $\rho_l=10^3$ kg/m^3, the liquid speed of sound $c_l=1.5\times10^3$ m/s, the liquid viscosity $\mu_l=10^{-3}$ N s/m^2, the gas adiabatic index $\gamma=1.4$, the surface tension coefficient $\sigma=0.07$ N/m, and the width of bubble layer $L=0.1$ m.

method is accurate and so is the computer code that implements it.

As the amplitude of the incident wave increases, nonlinear effects eventually lead to shock formation as demonstrated in Fig. 3, which shows the normalized transmitted (upper panel) and reflected waves at steady state for $P_{\rm inc}/P_\infty=0.01$, 0.1, 0.2 (dotted line), 0.3, 0.5, and 0.9; these waveforms are shown during the 20th cycle, by which time steady state has been reached. The results for $P_{\rm inc}/P_\infty=0.2$ are singled out using a dotted line because (38) gives a value of 0.18 for the shock-wave threshold, which is seen to be in good agreement with the numerical results. At the lowest drive the wave is essentially completely transmitted: the reflected component is very small and almost entirely consisting of the second harmonic. Indeed, the linear theory results (23) and (24) predict 100% transmission and zero reflection for these conditions. Figure 4 is the pressure distribution in the layer at time $20\times2\pi/\omega$ for all six driving amplitudes; again, the line corresponding to $P_{\rm inc}/P_\infty=0.2$ is dotted and shows the incipience of shock formation.

The analog of Fig. 2, but at the much higher amplitude

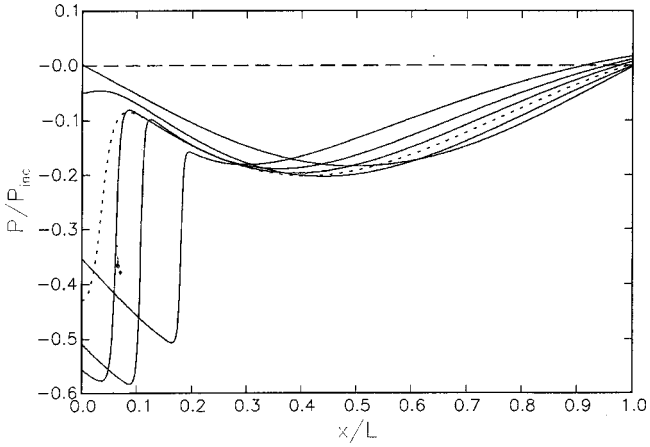


FIG. 4. Pressure field inside the bubble layer at time $20\times\omega/2\pi$ for different amplitudes of the incident wave, $P_{\rm inc}/P_\infty=0.01$, 0.1, 0.2 (dotted), 0.3, 0.5, 0.9 according to the quasi-equilibrium model. The frequency of the incident wave is equal to the first linear eigenfrequency of the layer $\omega=\omega_1$, with $\omega_1/2\pi=1.347$ kHz. The gas volume fraction is $\beta_0=0.133\%$; other conditions as specified at the beginning of Sec. V.

$P_{\rm inc}/P_\infty=0.7$, is shown in Fig. 5. The circles and squares connected by dashed lines are the numerical results while the solid and dotted lines show the linear theory predictions. Now that the threshold value is far exceeded, nonlinear effects are dominant, and the discrepancy between the two sets of results is very pronounced, especially when the incident frequency is close to the first linear eigenfrequency of the layer, $\omega/\omega_1\sim1$. The maximum of the resonance curve of the transmitted wave is about 0.6, rather than 1, and it occurs at a value of ω/ω_1 shifted to the left of the linear resonance condition, which indicates the expected softening behavior of the nonlinear oscillator. The considerable dissipation seen in this case is a consequence of shock formation in the layer.

The transient process that culminates in the formation of

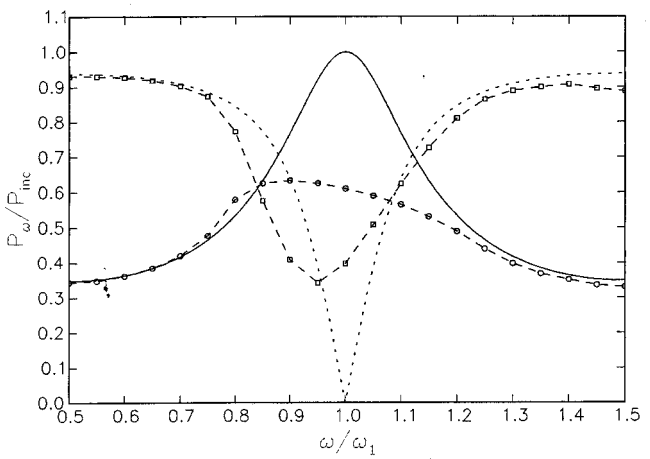


FIG. 5. Pressure amplitudes of the component at the incident wave frequency ω for the transmitted (circles) and reflected (squares) waves as functions of the incident frequency ω normalized by the first linear eigenfrequency of the layer according to the quasi-equilibrium model, $\omega_1/2\pi=1.347$ kHz. The dashed lines connecting circles and squares are only guides to the eye. The solid and dotted lines are the linear results given in Eqs. (23) and (24). The incident wave amplitude $P_{\rm inc}/P_\infty=0.7$ is greater than the threshold value for shock formation given by (38). The gas volume fraction is $\beta_0=0.133\%$; other conditions as specified at the beginning of Sec. V.

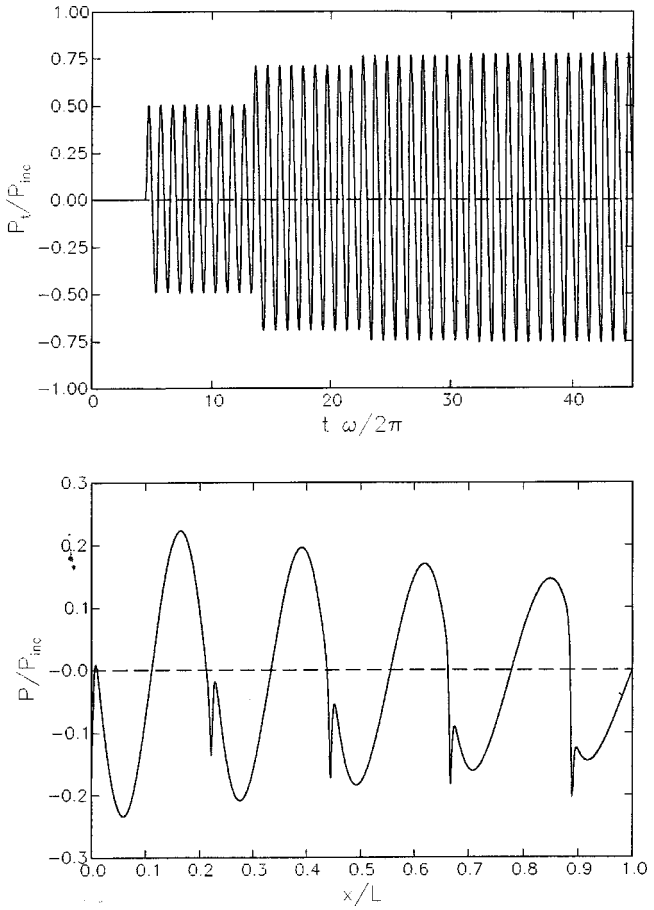


FIG. 6. Transmitted wave (upper panel) and pressure field inside the bubble layer at time $t=45\times 2\pi/\omega$ for $P_{\rm inc}/P_\infty=0.05$, $\beta_0=0.133\%$, and $\omega=9\omega_1=2\pi\times 12.1$ kHz according to the quasi-equilibrium model.

a shock is due to the gradual build-up of the multiple reflections that occur at the layer boundaries and is demonstrated in Fig. 6. The upper panel shows, as a function of the normalized time $\omega t/2\pi$, the normalized transmitted wave, which is seen to be progressively reinforced each time the shock developing inside the bubbly layer reflects at the layer boundary at $x=L$. Here the frequency of the incident wave is equal to the ninth linear eigenfrequency of the layer, $\omega/\omega_1=9$, and the polytropic index equals 1. The incident wave has amplitude $P_{\rm inc}/P_\infty=0.05$ and thus exceeds the threshold value (38) which, for these conditions, is 0.02. The time needed for the wave to travel from the left boundary $x=0$ to $x=L$ and back is about $9\times 2\pi/\omega$, and this is the approximate time separation of the steps that are clearly seen in the first few reflections. The lower panel of the figure shows the pressure field inside the layer at the end of the calculation shown in the upper panel, $45\times 2\pi/\omega$. The steady-state shape of the transmitted wave is shown in Fig. 7, with a clearly evident shock structure as expected.

The results of Figs. 6 and 7 should be compared with those presented in Fig. 3 of Druzhinin *et al.* (1996). The numerical scheme used in that work introduced numerical dispersion due to the discretization, which led to oscillations near the shock. By testing their code we realized that these oscillations are sensitive to the spatial step used in the calculation, and must therefore be considered an artifact of the numerical method used.

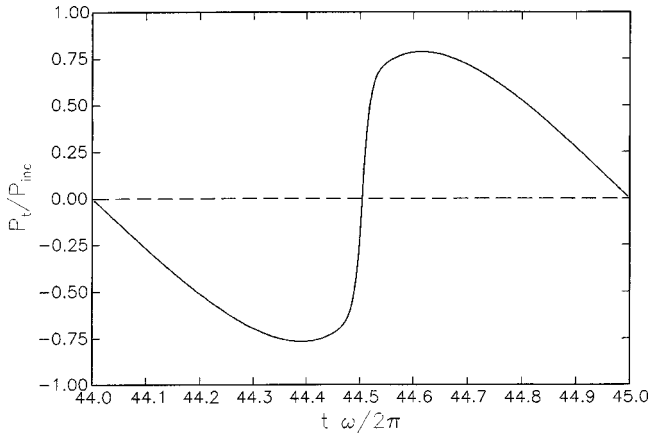


FIG. 7. Steady-state shape of the transmitted wave for the case of Fig. 6 ($P_{\rm inc}/P_\infty=0.05, \beta_0=0.133\%, \omega=9\omega_1$) according to the quasi-equilibrium model.

B. Complete model

We now turn to the complete model in which the bubble behavior is described by Eqs. (5)–(7). The inertia affecting the bubble pulsations now causes a strong dependence of the pressure wave upon the bubble radius, and thermal effects contribute to the damping of the wave. Thus, it is interesting to compare the predictions of this model with those of the polytropic approximation (which includes inertia but no thermal damping) and of the quasi-equilibrium model (which includes neither).

In choosing a suitable value of the polytropic index for use with the polytropic model we face the usual problem of the lack of a basis for this choice when the gas is neither clearly adiabatic nor isothermal. In our calculations we use for κ_p the linear-theory value κ_l defined in (17). Note that κ_l and ω_0 as given by the linearized theory of Sec. (III) depend on the driving frequency; the values that we quote below are calculated for the driving frequency of the incident wave ω.

In order to illustrate the effect of the bubble radius we consider the response of a layer driven near its first mode for two cases, both with the same gas volume fraction, $\beta_0=0.133\%$, but with a different radius of the constituent bubbles, $R_0=0.121$ and 1.21 mm. Again, the isothermal linear frequency of the lowest layer mode, according to (26), is $\omega_1/2\pi=1.347$ kHz.

In the first case the bubble equilibrium radius is $R_0=0.121$ mm; the linear bubble resonance frequency, 23.27 kHz, is thus much greater than that of the layer mode. With $\omega=\omega_1$ the linear polytropic index of the bubbles as given by (17) is $\kappa_l=1.035$, which confirms the essentially isothermal behavior of the gas.

Figure 8 shows the transmitted wave at steady state normalized by the amplitude of the incident wave as a function of the dimensionless time $t\omega/2\pi$ for two incident amplitudes, $P_{\rm inc}/P_\infty=0.01$ (upper panel) and $P_{\rm inc}/P_\infty=0.7$ (lower panel). The solid lines are the results of the complete model, while the dotted and dashed lines are the results for the polytropic and quasi-equilibrium models with $\kappa_p=1$, respec-

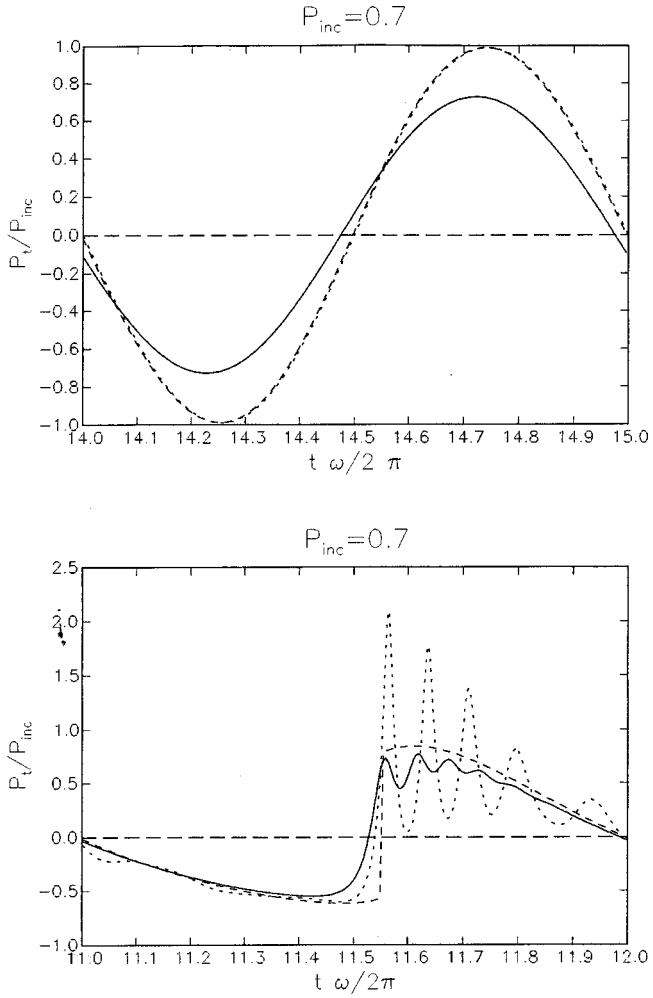


FIG. 8. Steady-state shape of the transmitted wave according to the complete model (solid line), compared with polytropic (dotted line) and quasi-equilibrium (dashed line) models with $\kappa_p=1$; upper panel $P_{\rm inc}/P_\infty=0.01$, lower panel $P_{\rm inc}/P_\infty=0.7$; $\omega/2\pi=1.347$ kHz, $\beta_0=0.133\%$, $R_0=0.121$ mm.

tively. When the incident amplitude is small (upper panel) the polytropic and quasi-equilibrium models give essentially the same result, which is expected in this case in which the incident frequency is much smaller than the bubble resonance frequency. The complete model predicts a somewhat smaller amplitude and a small phase shift, both due to the inclusion of thermal dissipation in the bubble motion, but behaves otherwise very similarly. In principle, these features contain information about the bubble size and possibly other quantities, although its extraction might be problematic in practice.

For the larger-amplitude excitation (lower panel of Fig. 8) all three models predict shock formation, but the differences among them are more pronounced. The complete model (solid line) shows a transmitted wave with slight oscillations near the maximum, while these oscillations are highly exaggerated by the polytropic model (dotted line). The same qualitative difference is encountered when the two models are applied to shock waves in bubbly liquids (Watanabe and Prosperetti, 1994). Since these oscillations are a consequence of bubble inertia, their absence in the quasi-equilibrium model (dashed line) is not surprising. The reason

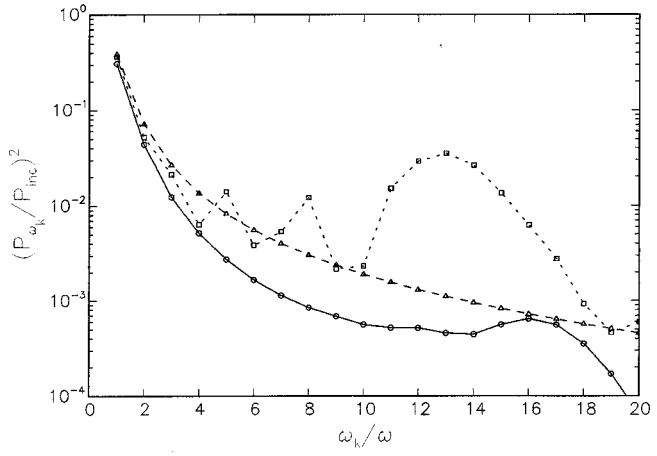


FIG. 9. Power spectrum of the transmitted wave for the case of the lower panel of Fig. 8 according to the complete model (circles), compared with polytropic (squares) and quasi-equilibrium (triangles) models with $\kappa_p=1$; $P_{\rm inc}/P_\infty=0.7$, $\omega/2\pi=1.347$ kHz, $R_0=0.121$ mm. The lines are only guides to the eye.

why this feature is encountered at this higher amplitude but not at the lower amplitude of the upper panel of Fig. 8 is that the formation of the shock introduces a much shorter characteristic time scale in the wave, which is not too far from the resonant period of the bubbles. Due to the absence of thermal effects, the polytropic model is less dissipative than the complete one, and the shock time scale is accordingly shorter: the smaller damping and the shorter time scale combine to cause the prominent oscillations of the result. If these oscillations are averaged out in the mind's eye, one sees a substantial similarity among the different profiles which is due to the fact that the underlying, relatively slowly varying, wave structure is only slightly damped in all models. In any event, it may be noted that the high-frequency oscillations of the complete model are so strongly damped that the quasi-equilibrium model ends up being a better approximation to the actual behavior than the polytropic model.

The power spectrum of the transmitted wave for $P_{\rm inc}/P_\infty=0.7$ is shown in Fig. 9. The circles represent the complete model, the squares the polytropic model, and the triangles the quasi-equilibrium model; the lines connect the symbols as an aid to the eye. The more dissipative nature of the complete model is evident from this comparison, as expected. The polytropic model exhibits several peaks corresponding to the harmonics of the drive and reflecting the strong oscillations of the lower panel of Fig. 8; these features are caused by the strongly nonlinear bubble response. The complete model, instead, only shows a mild resonance in correspondence with the bubble fundamental resonance at $\omega/\omega_1\simeq 16$.

In order to illustrate the behavior of the layer when the bubble natural frequency is not as different from the layer modal frequency as in the previous case, we consider now bubbles with an equilibrium radius of 1.21 mm, for which $\omega_0/2\pi=2.647$ kHz and $\kappa_l=1.35$. In this case the bubble natural frequency is about twice that of the lowest eigenmode of the layer.

Figure 10 shows the components of the transmitted (upper panel) and reflected (lower panel) waves at the incident

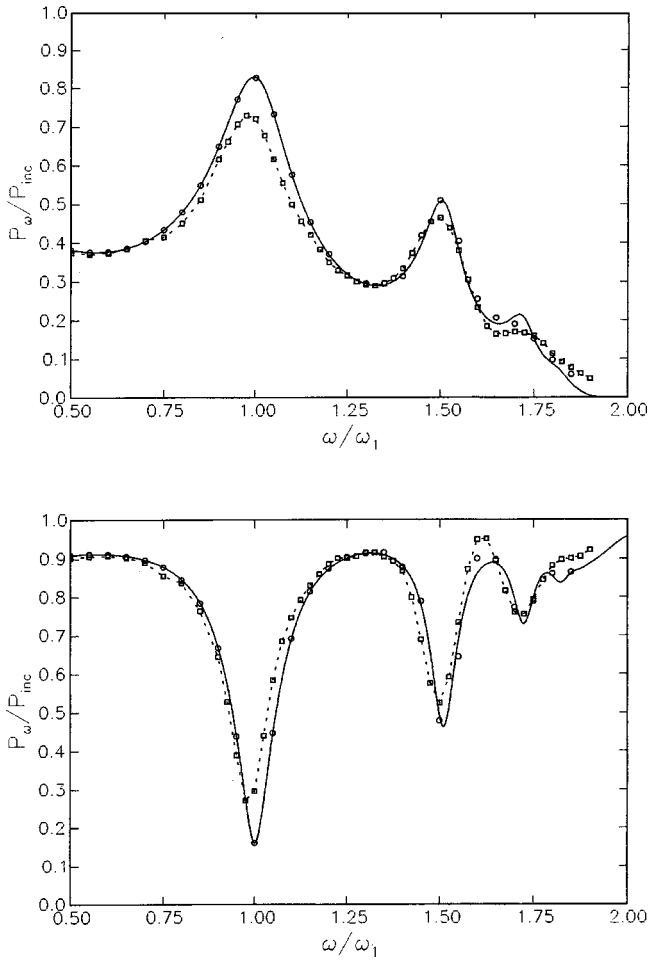


FIG. 10. Pressure amplitudes of the component at the incident wave frequency ω for the transmitted (upper panel) and reflected (lower panel) waves as functions of the incident frequency ω normalized by the first quasi-equilibrium linear eigenfrequency of the layer, $\omega_1/2\pi=1.347$ kHz. The circles and squares are the numerical results of the complete model for amplitudes $P_{\rm inc}/P_\infty=0.05$ and $P_{\rm inc}/P_\infty=0.7$, respectively. The solid curves are the linear results given by (21) and (22). The dotted lines connecting squares are only guides to the eye. The bubble radius is 1.21 mm.

frequency for $P_{\rm inc}/P_\infty=0.05$ (circles) and $P_{\rm inc}/P_\infty=0.7$ (squares) according to the complete model as functions of the normalized incident frequency ω/ω_1, with $\omega_1/2\pi=1.347$ kHz as before; the dotted lines connect the symbols as an aid to the eye. The lower-amplitude results match very well the analytic linear results of (21) shown by the solid line. Remarkably, when normalized by the incident pressure as here, the higher-amplitude results are only slightly different, with somewhat less pronounced maxima and minima and a slight shift to lower frequencies. This is surprising in view of the earlier results shown in Fig. 5, where a much greater difference was encountered. The explanation lies in the fact that shock formation in that case caused a strong energy dissipation that is absent in the complete-model results of Fig. 10.

The frequency range in Fig. 10 is near the bubble resonance where, for equal bubbles, the phase speed of pressure waves decreases substantially (see, e.g., Commander and Prosperetti, 1989). This circumstance has the effect of moving all the layer resonances to lower frequencies, the more

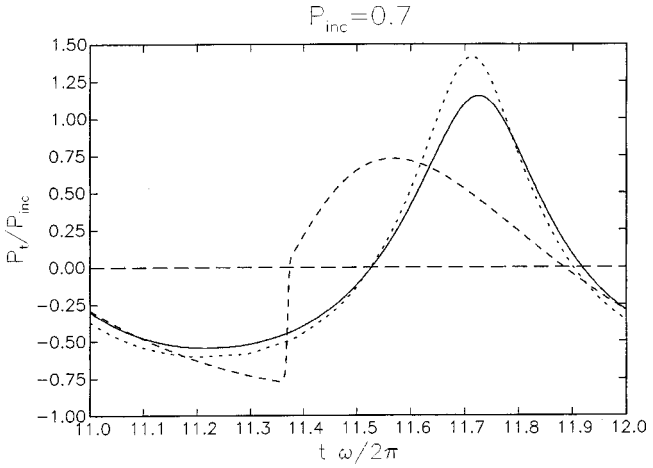


FIG. 11. Steady-state waveform of the transmitted wave for $P_{\rm inc}/P_\infty=0.7$ and $\omega/\omega_1=0.975$ according to the complete model (solid line), compared with polytropic (dotted line) and quasi-equilibrium (dashed line) models with $\kappa_p=1.35$; all conditions as in Fig. 10.

the closer they are to the bubble natural frequency: the decreasing spacing between maxima and minima with increasing frequency is evident from the figure. In the quasi-equilibrium model, on the other hand, the absence of dispersion gives an equal spacing between maxima and minima.

The transmitted waveforms near the first resonance, $\omega/\omega_1=0.975$, are shown in Fig. 11 for an incident wave amplitude $P_{\rm inc}/P_\infty=0.7$. In this case, in which the bubbles behave nearly adiabatically, the total damping is small and the complete (solid line) and polytropic (dotted line) model results are, accordingly, close; the dashed line is the quasi-equilibrium model, which shows an unrealistic shock structure that is eliminated by inertia in the other models.

The power spectrum of the transmitted wave is shown in Fig. 12. As compared with the previous case of Fig. 8, the spectrum for the complete model (circles) exhibits a much faster decay due to the fact that all these modes are above the fundamental bubble resonance; the polytropic result (squares) is similar, confirming that thermal damping effects

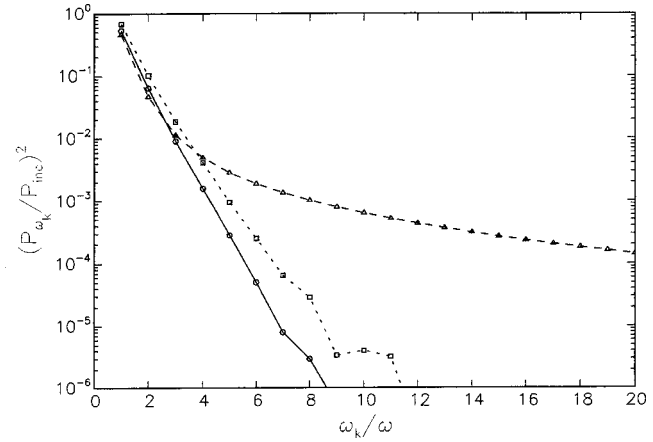


FIG. 12. Power spectrum of the transmitted wave for the case of Fig. 11 according to the complete model (circles), compared with polytropic (squares) and quasi-equilibrium (triangles) models with $\kappa_p=1.35$. The lines are only guides to the eye.

are small, while the quasi-equilibrium model spectrum (triangles) decays much more slowly due to the presence of the shock.

VI. RESULTS: DUAL-FREQUENCY EXCITATION

We now consider the behavior of the bubble layer when the incident wave contains two frequencies, $\omega^{(1)}$ and $\omega^{(2)}$. The nonlinearity will then produce several Fourier components and in particular the difference frequency $\Omega=\omega^{(2)}-\omega^{(1)}$. By arranging for both incident frequencies and the difference frequency to correspond to normal modes of the layer, an effective generation of the low-frequency component may be expected to take place.

In order to avoid the strong damping that has plagued earlier attempts to use bubbly liquids in parametric arrays, it is important to operate under conditions in which the bubble natural frequency is higher than either incident frequency. As noted before in Sec. II, this condition corresponds to the assumptions that justify the quasi-equilibrium model of Eq. (9) with $\kappa_p=1$ and, for this reason, we start the study of the low-frequency signal generation with this model. Except where explicitly noted, the parameters used in the calculations have the same value as given at the beginning of Sec. V.

We consider a bubble layer excited by a biharmonic incident wave with components at frequencies $\omega^{(1)}$ and $\omega^{(2)}$ close to the ninth and tenth eigenmodes of the layer; for all cases considered in this section both incident components have the same amplitude $P_{\rm inc}$, $A^2=31$, and $\kappa_p=1$. In order to operate much below the bubble resonance frequency, we increase the thickness of the layer to $L=0.4$ m; for isothermal conditions the frequency of the fundamental model is $\omega_1/2\pi=0.337$ kHz.

A typical example of the steady-state shape of the transmitted wave is shown in the upper panel of Fig. 13 where $\omega^{(1)}=9\omega_1=2\pi\times3.03$ kHz and $\omega^{(2)}=10\omega_1=2\pi\times3.37$ kHz, so that the difference frequency equals the first eigenfrequency of the layer, $\omega_1=0.337$ kHz, providing the resonance conditions for the low-frequency signal. The amplitude of each one of the incident components is $P_{\rm inc}/P_\infty=0.5$. In addition to the nonlinearly generated low-frequency component, the slow modulation of the waveform reflects the standard beats due to the superposition of two oscillations with close periods (the same phenomenon is visible in Fig. 17 below). The pressure field inside the bubble layer at the times $t=8\times2\pi/\omega_1$ and $t=8.5\times2\pi/\omega_1$ is shown in the lower panel of Fig. 13. Since the amplitude of the incident wave is much larger than the threshold value for shock formation (equal to 0.02 for this case), shocks form in the layer and are reflected at the boundaries, which is responsible for the characteristic double shock-front structure present in the pressure field in the lower panel of Fig. 13. The internal shocks confer a sawtooth structure to the transmitted wave as shown in the upper panel of Fig. 13. The relative power of the difference-frequency harmonic in the transmitted wave, i.e., $(P_\Omega/P_{\rm inc})^2$, is 1.12×10^{-2}.

The dependence of the low-frequency output on the incident frequencies is shown in Fig. 14 where the pump frequencies $\omega^{(1)}$ and $\omega^{(2)}$ are changed in such a way that the difference-frequency is fixed and equal to the layer first eigenfrequency, $\Omega=\omega^{(2)}-\omega^{(1)}=\omega_1$. The circles, squares,

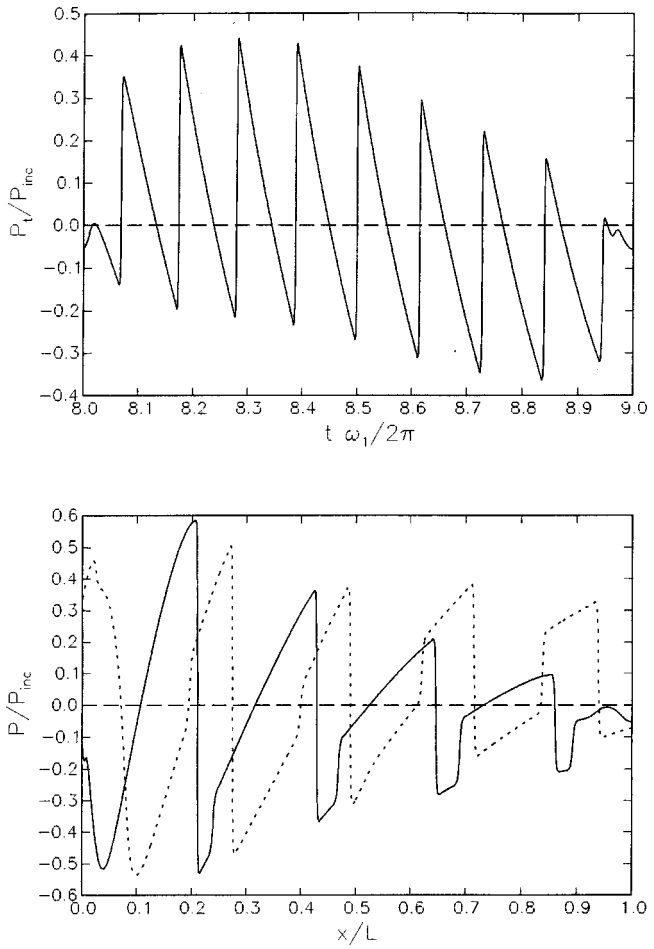


FIG. 13. Steady-state shape of transmitted wave (upper panel) produced by the quasi-equilibrium model and pressure field inside the bubble layer at time $t=8\times2\pi/\omega_1$ (solid line) and $t=8.5\times2\pi/\omega_1$ (dotted line) for dual-frequency excitation with $\omega^{(1)}=9\omega_1=2\pi\times3.03$ kHz, $\omega^{(2)}=10\omega_1=2\pi\times3.37$ kHz and amplitude $P_{\rm inc}/P_\infty=0.5$.

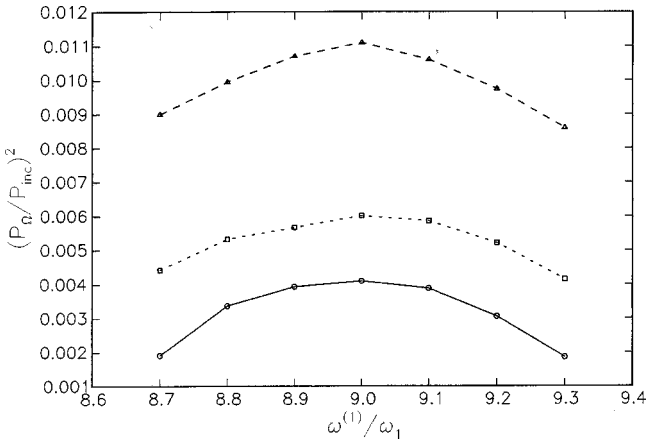


FIG. 14. Relative power of the low-frequency component in the transmitted wave according to the quasi-equilibrium model for dual-frequency excitation as a function of $\omega^{(1)}$ with $\omega^{(2)}=\omega^{(1)}+\omega_1$. The circles, squares, and triangles are the results for $P_{\rm inc}/P_\infty=0.1$, 0.2, and 0.5 respectively. The lines are only guides to the eye.

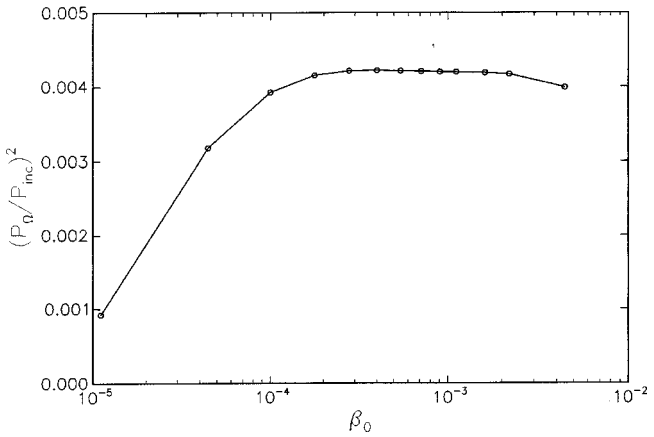


FIG. 15. Relative power of the low-frequency component predicted by the quasi-equilibrium model for dual-frequency excitation as a function of the gas volume fraction in the layer; $\omega^{(1)}=9\omega_1=2\pi\times3.03$ kHz, $\omega^{(2)}=10\omega_1=2\pi\times3.37$ kHz, and $P_{inc}/P_\infty=0.1$.

and triangles are the results for $P_{inc}/P_\infty=0.1$, 0.2, and 0.5, respectively. The lines are only guides to the eye. As expected, the relative power of the low-frequency signal has a maximum around $\omega^{(1)}=9\omega_1$ when both pump frequencies are at resonance conditions.

Figure 15 shows the relative low-frequency transmitted power as a function of the parameter β_0 for $P_{inc}/P_\infty=0.1$, again for the quasi-equilibrium model. With increasing gas volume fraction, the amplitude of the difference-frequency transmitted wave quickly saturates to an approximately constant value, after which it begins to decline. We have found that, in order to obtain converged results, the number of grid points had to be increased to 6400 at the higher volume fractions. For example, the difference between the values of $(P_\Omega/P_{inc})^2$ as computed with 3200 or 6400 points is 2.4% for $\beta=1.11\times10^{-3}$, and increases to 5.8% for $\beta=2.84\times10^{-3}$.

Figure 16 shows the relative power of the low-frequency signal $(P_\Omega/P_{inc})^2$ of the transmitted (circles) and reflected (squares) waves as a function of the amplitude of the incident wave components P_{inc}/P_∞ when $\omega^{(1)}=9\omega_1$ and $\omega^{(2)}=10\omega_1$; the solid and dotted lines are for the complete model while the dot-dash and dashed lines are for the quasi-equilibrium model. Here the bubble radius is $R_0=50$ μm, with a linear resonance frequency $\omega_0/2\pi=55.86$ kHz and $\kappa_l=1.01$.

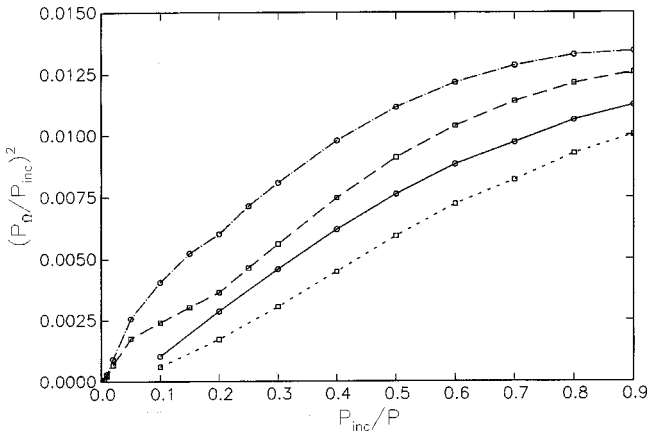


FIG. 16. Relative power of the low-frequency component in the transmitted (circles) and reflected (squares) waves for dual-frequency excitation as a function of the amplitude P_{inc}/P_∞ of the incident wave components for $\omega^{(1)}=9\omega_1=2\pi\times3.03$ kHz and $\omega^{(2)}=10\omega_1=2\pi\times3.37$ kHz. The lines are only guides to the eye. The solid and dotted lines are for the complete model while the dot-dash and dashed lines are for the quasi-equilibrium model with $\kappa_p=1$. The bubble radius is 50 μm.

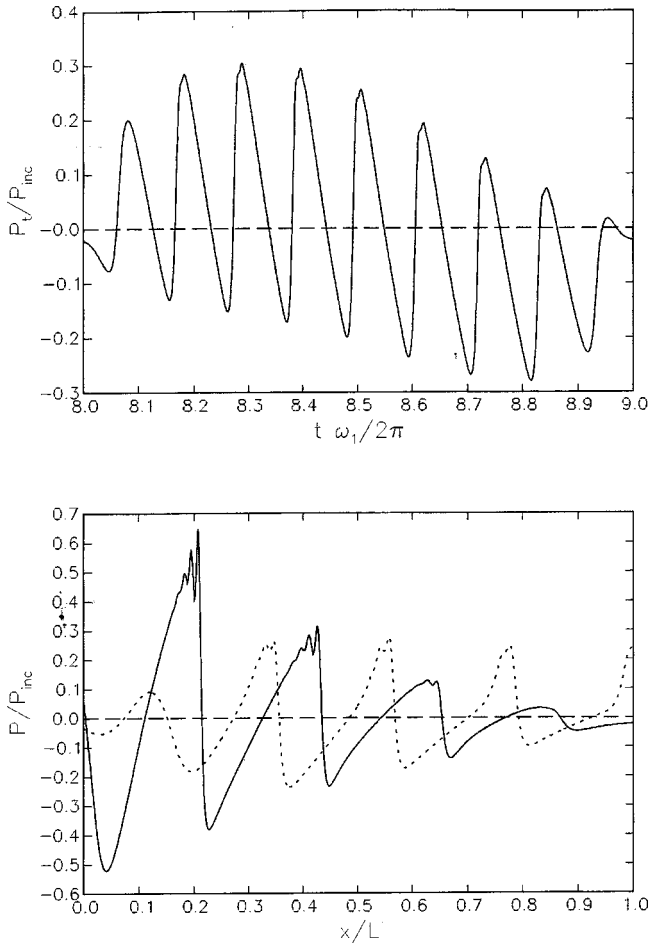


FIG. 17. Steady-state shape of transmitted wave (upper panel) produced by the complete model and pressure field inside the bubble layer at times $t=8\times2\pi/\omega_1$ and $t=8.5\times2\pi/\omega_1$ for dual-frequency excitation with $\omega^{(1)}=9\omega_1=2\pi\times3.03$ kHz, $\omega^{(2)}=10\omega_1=2\pi\times3.37$ kHz, and amplitude $P_{inc}/P_\infty=0.5$. These results should be compared with those given by the quasi-equilibrium model presented in Fig. 13.

The relative power of the low-frequency signal increases with the amplitude of the incident wave, but the proportionality to the square of the incident wave amplitude predicted by the weakly nonlinear theory (Druzhinin *et al.*, 1996) is found only for small amplitudes, $P_{inc}/P_\infty\leqslant0.01$, and is not apparent from this figure. The quasi-equilibrium-model results of Fig. 16 are much lower than the numerical results presented in Figs. 6(a) and (b) of Druzhinin *et al.* (1996), apparently once again due to numerical artifacts of that work.

Figure 17 shows the steady-state temporal transmitted waveform (upper panel) and the pressure field inside the bubble layer at time $t=8\times2\pi/\omega_1$ for $\omega^{(1)}=9\omega_1$, $\omega^{(2)}=10\omega_1$, $P_{inc}/P_\infty=0.5$, for 50-μm-radius bubbles. This figure should be compared with Fig. 13 for the quasi-equilibrium model. The transmitted wave still has a sawtooth

TABLE I. Relative power of the transmitted and reflected waves produced by the complete model for a dual-frequency incident wave with $\omega^{(1)}=9\omega_1$, $\omega^{(2)}=10\omega_1$, $P_{\rm inc}/P_\infty=0.1$; other conditions as given at the beginning, of Sec. V except that the layer thickness is 0.4 m ($\omega_1/2\pi=337$ Hz) and 0.1 m ($\omega_1/2\pi=1{,}347$ Hz).

R_0 (mm)	$\omega_0/2\pi$ (Hz)	κ_l	$\omega_1/2\pi$ (Hz)	(Transmitted power)/$\lvert P_{\rm inc}\rvert^2$		(Reflected power)/$\lvert P_{\rm inc}\rvert^2$	
				Total	At Ω	Total	At Ω
0.05	55,86	1.01	337.0	0.16	1.0×10^{-3}	0.77	6.1×10^{-4}
0.05	55,90	1.08	1,347	0.02	2.5×10^{-4}	0.95	1.7×10^{-4}
0.1	28,76	1.08	337.0	0.02	2.4×10^{-4}	0.97	1.6×10^{-4}
0.1	30,57	1.22	1,347	0.02	2.4×10^{-4}	0.94	1.6×10^{-4}
1	3,22	1.36	337.0	2.4×10^{-5}	2.4×10^{-5}	1.72	2.0×10^{-5}
1	3,24	1.38	1,347	0.20	2.5×10^{-6}	1.79	6.3×10^{-6}

appearance, but the peaks are less sharp, which is evidence of the stronger damping affecting the higher frequencies. Due to the increased dissipation, the shocks inside the layer are strongly damped before they reach the right end and their reflection there is thus less strong than before, which explains the marked differences between the lower panels of Figs. 17 and 13. Now oscillations near the back of the shock are present; these are related, as before, to bubble inertia. The relative power of the difference-frequency harmonic in the transmitted wave is 7.6×10^{-3}; for the same conditions, but at $P_{\rm inc}/P_\infty=0.1$, the relative power is 1.0×10^{-3}. The corresponding results for the quasi-equilibrium model are 11.2×10^{-3} and 4.05×10^{-3}, respectively, for $P_{\rm inc}/P_\infty=0.5$ and 0.1. In both cases the relative power of the transmitted difference-frequency predicted by the complete model is smaller than that of the quasi-equilibrium model.

The present problem contains a large number of parameters, and it is not practical to present an exhaustive investigation of the entire parameter space. Some further insight into the dependence of the low-frequency transmitted component on various quantities can be gained from Table I which summarizes the results of several calculations for different parameters of the layer and air bubbles. In all cases $A=\sqrt{31}$, and the layer is excited by a biharmonic wave at frequencies $\omega^{(1)}=9\omega_1$ and $\omega^{(2)}=10\omega_1$ with amplitude $P_{\rm inc}/P_\infty=0.1$. The bubble layer has a thickness of 40 cm ($\omega_1/2\pi=337$ Hz) and 10 cm ($\omega_1/2\pi=1{,}347$ Hz); $\kappa_p=\kappa_l$ is evaluated at $\omega^{(2)}$. These results show that the relative power of the low-frequency signal is larger when the frequencies of the incident wave are much smaller than the bubble resonance frequency and the bubble oscillations are isothermal. These conditions might suggest the use of the simpler quasi-equilibrium model; however, in all cases, we have found that the latter tends to overpredict the amplitude of the low-frequency signal.

VII. CONCLUSIONS

The results presented in this paper correct and expand the scope of the earlier ones of Druzhinin *et al.* (1996). In comparison with that work, we have used a more realistic description of the bubble behavior which includes the effects of radial inertia, with the associated dispersion, and of the gas thermal behavior, with the attendant energy losses. As expected, several details of the predictions of the earlier model are modified by these effects, although the basic character remains. Thus, we find a propensity for shock-wave generation in the bubbly layer which gives rise to a transmitted wave with a sawtooth character.

We have also studied the possibility of enhanced low-frequency difference-wave generation through the exploitation of the resonances of the bubbly layer. We have concluded that the estimates presented in our earlier work were excessively large due to a combination of the idealizations of the model and an insufficiently accurate numerical scheme. However, even after correction, we find that a difference-wave power of the order of 1% of the incident power (Fig. 16) can be generated using incident wave amplitudes of less than 100 kPa. Thus, this technique for the parametric generation of low-frequency waves may have practical value.

An important aspect of the phenomenon, confirmed by this study, is that operation near the resonance frequency of the individual bubbles is detrimental to the energy conversion efficiency due to the strong dissipation of the bubble motion in this frequency range.

In order to avoid the practical difficulties connected with the generation and control of suitable bubbles, it might be expedient to apply the principle described in this paper to other systems. For example, one would expect similar phenomena to occur in porous waterlike (or rubberlike) media in which the shear modulus is small and plays the role of gas compressibility in bubbles; some experiments of this type are reported in Belyaeva and Timanin (1991). Like bubbles, pores in such media provide very strong nonlinearity (see, e.g., Naugolnykh and Ostrovsky, 1998, Sec. 1.4). At the same time, it is much easier to have small and almost equal-size pores, the system is more stable, and losses are typically smaller.

ACKNOWLEDGMENTS

The authors express their gratitude to the Office of Naval Research for the support of this work.

Akhatov, I., Parlitz, U., and Lauterborn, W. (**1994**). ''Pattern formation in acoustic cavitation,'' J. Acoust. Soc. Am. **96**, 3627–3635.

Belyaeva, I., and Timanin, E. (**1991**). ''Experimental study of the nonlinear properties of porous elastic media,'' Sov. Phys. Acoust. **37**, 533–544.

Caflisch, R. E., Miksis, M. J., Papanicolaou, G., and Ting, L. (**1985**). ''Effective equations for wave propagation in bubbly liquids,'' J. Fluid Mech. **153**, 259–273.

Carstensen, E. L., and Foldy, L. L. (**1947**). ''Propagation of sound through a liquid containing bubbles,'' J. Acoust. Soc. Am. **19**, 481–501.

Clay, C. S., and Medwin, H. (**1977**). *Acoustical Oceanography* (Wiley, New York).
Colonius, T., d'Auria, F., and Brennen, C. E. (**2000**). ''Acoustic saturation in bubbly cavitating flow adjacent to an oscillating wall,'' Phys. Fluids **12**, 2752–2761.
Commander, K. W., and Prosperetti, A. (**1989**). ''Linear pressure waves in bubbly liquids: comparison between theory and experiments,'' J. Acoust. Soc. Am. **85**, 732–746.
d'Agostino, L., and Brennen, C. E. (**1988**). ''Acoustical absorption and scattering cross-sections of spherical bubble clouds,'' J. Acoust. Soc. Am. **84**, 2126–2134.
Druzhinin, O. A., Ostrovsky, L. A., and Prosperetti, A. (**1996**). ''Low-frequency acoustic wave generation in a resonant bubble layer,'' J. Acoust. Soc. Am. **100**, 3570–3580.
Feng, Z. C., and Leal, L. G. (**1997**). ''Nonlinear bubble dynamics,'' Annu. Rev. Fluid Mech. **29**, 201–243.
Gasenko, V. G., Nakoryakov, V. E., and Shreiber, I. R. (**1979**). ''Nonlinear disturbances in a liquid containing gas bubbles,'' Sov. Phys. Acoust. **25**, 385–388.
Harten, A. (**1983**). ''High resolution schemes for hyperbolic conservation laws,'' J. Comput. Phys. **49**, 357–393.
Kamath, V., and Prosperetti, A. (**1989**). ''Numerical integration methods in gas-bubble dynamics,'' J. Acoust. Soc. Am. **85**, 1538–1548.
Kamath, V., Prosperetti, A., and Egolfopoulos, F. (**1993**). ''A theoretical study of sonoluminescence,'' J. Acoust. Soc. Am. **94**, 248–260.
Kogarko, B. S. (**1964**). ''One-dimensional unsteady motion of a liquid with an initiation and progression of cavitation,'' Dokl. Akad. Nauk SSSR **155**, 779–782 (in Russian).
Kustov, L. M., Nazarov, V. E., Ostrovsky, L. A., Sutin, A. M., and Zamolin, S. V. (**1982**). ''Parametric acoustic radiation with a bubble layer,'' Acoust. Lett. **6**, 15–17.
Kuznetsov, V. V., Nakoryakov, V. E., Pokusaev, G., and Shreiber, I. R. (**1978**). ''Propagation of perturbations in a gas-liquid mixture,'' J. Fluid Mech. **85**, 85–96.
Medwin, H., and Clay, C. S. (**1997**). *Fundamentals of Ocean Acoustics* (Morgan Kauffman, San Francisco).
Naugolnykh, K., and Ostrovsky, L. (**1998**). *Nonlinear Wave Processes in Acoustics* (Cambridge U.P., Cambridge, UK), Chap. 1, pp. 25–30.
Nigmatulin, R. I. (**1991**). *Dynamics of Multiphase Media* (Hemisphere, Washington), Vol. 1, Chap. 1, pp. 98–113.
Omta, R. (**1987**). ''Oscillations of a cloud of small bubbles of small and not so small amplitude,'' J. Acoust. Soc. Am. **82**, 1018–1033.
Ostrovsky, L. A., Sutin, A. M., Soustova, I. A., and Matveyev, A. I. (**1998**). ''Nonlinear, low-frequency sound generation in a bubble layer: Theory and laboratory experiments,'' J. Acoust. Soc. Am. **104**, 722–726.
Pierce, A. D. (**1989**). *Acoustics* (Acoustical Society of America, Woodbury, NY), Chap. 3, pp. 130–140.
Plesset, M. S., and Prosperetti, A. (**1977**). ''Bubble dynamics and cavitation,'' Annu. Rev. Fluid Mech. **9**, 145–185.
Prosperetti, A. (**1984**). ''Bubble phenomena in sound fields: Part one,'' Ultrasonics **22**, 69–77.
Prosperetti, A. (**1991**). ''The thermal behaviour of oscillating gas bubbles,'' J. Fluid Mech. **222**, 587–616.
Prosperetti, A. (**2001**). ''Fundamental acoustic properties of bubbly liquids,'' in *Handbook of Elastic Properties of Solids, Liquids, and Gases*, edited by M. Levy, H. E. Bass, and R. Stern (Academic, New York), Vol. 4, pp. 183–205.
Prosperetti, A., and Lezzi, A. (**1986**). ''Bubble dynamics in a compressible liquid. Part 1. First-order theory,'' J. Fluid Mech. **185**, 289–321.
Twersky, V. (**1962**). ''On scattering of waves by random distributions. I. Free-space scatterer formalism,'' J. Math. Phys. **3**, 700–715.
van Wijngaarden, L. (**1968**). ''On the equations of motion for mixtures of liquid and gas bubbles,'' J. Fluid Mech. **33**, 465–474.
Watanabe, M., and Prosperetti, A. (**1994**). ''Shock waves in dilute bubbly liquids,'' J. Fluid Mech. **274**, 349–381.
Waterman, P. C., and Truell, R. (**1961**). ''Multiple scattering of waves,'' J. Math. Phys. **2**, 512–537.
Zabolotskaya, E. A. (**1977**). ''Two self-action mechanisms for sound waves propagating in a gas-liquid mixture,'' Sov. Phys. Acoust. **23**, 338–340.
Zabolotskaya, E. A., and Soluyan, S. I. (**1973**). ''Emission of harmonic and combination frequency waves by air bubbles,'' Sov. Phys. Acoust. **18**, 396–398.
Zhang, D. Z., and Prosperetti, A. (**1994**). ''Ensemble phase-averaged equations for bubbly flows,'' Phys. Fluids **6**, 2956–2970.

An acoustic streaming instability in thermoacoustic devices utilizing jet pumps

S. Backhaus[a] and G. W. Swift
Condensed Matter and Thermal Physics Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 30 August 2002; accepted for publication 29 November 2002)

Thermoacoustic-Stirling hybrid engines and feedback pulse tube refrigerators can utilize jet pumps to suppress streaming that would otherwise cause large heat leaks and reduced efficiency. It is desirable to use jet pumps to suppress streaming because they do not introduce moving parts such as bellows or membranes. In most cases, this form of streaming suppression works reliably. However, in some cases, the streaming suppression has been found to be unstable. Using a simple model of the acoustics in the regenerators and jet pumps of these devices, a stability criterion is derived that predicts when jet pumps can reliably suppress streaming. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1543588]

PACS numbers: 43.25.Nm, 43.35.Ud [MFH]

I. INTRODUCTION

Recently, thermoacoustic-Stirling hybrid engines[1–4] and refrigerators[4,5] that utilize traveling-wave phasing have been investigated. These devices are composed of a sandwich of three heat exchangers embedded in a looped acoustic network. The heat exchangers include an ambient heat exchanger, a stack or regenerator, and a heat exchanger at the ''working'' temperature which is above or below ambient depending on whether the device is an engine or refrigerator. These devices are filled with a thermodynamic working fluid, typically a pressurized ideal gas. The looped acoustic network may have distributed impedances, i.e., may have propagation lengths on the order of an acoustic wavelength,[3] or it may have lumped elements, i.e., with lengths much shorter than 1/4 of an acoustic wavelength.[1,2,4,5] In this article, the focus will be on engines and refrigerators that utilize regenerators embedded in a lumped-element acoustic network, because an engine of this type has already demonstrated high efficiency and refrigerators of this type promise higher efficiencies than existing orifice pulse tube refrigerators (OPTRs) at noncryogenic temperatures. A schematic drawing of a general lumped-element device is shown in Fig. 1. These devices are referred to as thermoacoustic-Stirling hybrid engines (TASHEs) and feedback pulse tube refrigerators (FPTRs).

Both TASHEs and FPTRs rely on their lumped acoustic networks for two roles. First, the network allows acoustic power to feed back from the working-temperature end of the regenerator to the ambient end. In an OPTR, this power is dissipated in an acoustic resistance at room temperature, while the acoustic network of an FPTR recycles this power into the ambient end of the regenerator.[5] This reduces the input power requirement and increases the efficiency of an FPTR compared to an OPTR.[6] In a TASHE, the acoustic power circulating in the acoustic network takes the place of mechanical components in various types of Stirling engines, such as cranks, linkages, or pistons, reducing the mechanical complexity of the engine.[2] In both devices, the second role of the network is to set the magnitude and phase of the acoustic impedance in the regenerator. By shifting the phase of the velocity oscillation by $\approx 90°$, the network transforms the standing-wave phasing in the resonator into nearly traveling-wave phasing in the regenerator. Also, the network keeps the magnitude of the acoustic impedance high so that viscous losses in the regenerator are reduced to an acceptable level.[2,5]

Although the lumped acoustic network makes both the TASHE and FPTR possible, it also introduces a complication. The network is essentially a wide-open tube that connects one end of the regenerator to the other. This topology allows for acoustic streaming, a steady flow of mass, around the looped network and through the regenerator. In fact, the acoustic power circulating around the looped network inherently encourages such streaming[7] in the direction from the ambient heat exchanger, through the regenerator, and to the working heat exchanger. A steady mass flow, no matter its direction, causes a heat leak to (FPTR) or from (TASHE) the working heat exchanger, drastically lowering the efficiency of both the TASHE and FPTR.

To combat this, a membrane or bellows could be used to block streaming around the loop while permitting oscillatory flow, but this might compromise the reliability inherent in a device that would otherwise have no moving parts. Alternatively, a nonlinear device termed a jet pump can be used to generate a time-averaged pressure, imposing low pressure at the ambient end of the regenerator so as to oppose the inherent tendency to stream. Typically, the geometry of the jet pump can be adjusted until the absence of streaming is indicated by a nearly linear temperature distribution through the regenerator.[2,5] In all TASHEs[2,8–10] that we have constructed to date, this technique has worked reliably. In contrast, one FPTR[11] that we have constructed demonstrated peculiar behavior. In that FPTR, it was impossible to adjust the geometry of the jet pump to stably cancel the streaming mass flux. By observing the temperature distribution in the regenerator, it was clear that there was always vigorous acoustic stream-

[a]Electronic mail: backhaus@lanl.gov

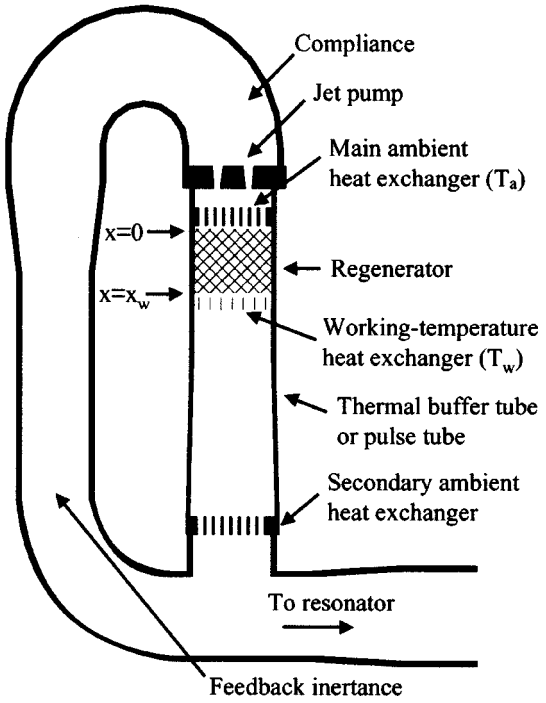


FIG. 1. Schematic drawing of a lumped-element FPTR or TASHE. An FPTR consumes acoustic power from the resonator and produces cooling at the working-temperature heat exchanger. Acoustic power that exits the working-temperature heat exchanger is recycled via the feedback loop to the main ambient heat exchanger. A TASHE supplies acoustic power to the resonator and circulates acoustic power around the feedback loop. As the power flows through the regenerator, it is amplified. Power in excess of that needed to maintain the power circulation in the feedback loop flows into the resonator to drive an acoustic load. Both the TASHE and FPTR are filled with a pressurized gas that serves as the thermodynamic working substance.

ing in one direction or the other. Also, the system demonstrated a hysteresis indicative of some type of instability. In another FPTR,[5] some operating points were stable while others were unstable.

The rest of this article presents a calculation that predicts when a TASHE or FPTR with a jet pump to suppress streaming is unstable to small perturbations. The few experimental observations described above are roughly consistent with these predictions. The temperature dependences of viscosity and density are key aspects of the calculation. If a small temperature change occurs in the heart of the regenerator, the viscosity and density will change and so the flow velocities through the regenerator may change. If the streaming mass flow changes in a direction that causes a temperature change in the heart of the regenerator of the same sign as the original change, positive feedback occurs and the streaming can grow catastrophically. The instabilities observed in some double-inlet pulse tube refrigerators[12,13] may also be due in part to this mechanism.

II. STEADY-STATE SOLUTION

The instability is investigated using a typical linear perturbation approach, where the acoustic variables are taken to be their equilibrium values plus a small perturbation. Using the usual acoustic expansion and time-harmonic notation,[14] the steady-state solution can be written

$$p(x,t)=p_m+\mathrm{Re}\,[p_1(x)e^{i\omega t}]+p_{2,0}(x), \tag{1}$$

$$U(x,t)=\mathrm{Re}\,[U_1(x)e^{i\omega t}]+U_{2,0}(x), \tag{2}$$

$$T(x,t)=T_m(x)+\mathrm{Re}\,[T_1(x)e^{i\omega t}], \tag{3}$$

where p, U, and T are the gas pressure, volumetric flow rate, and laterally spatially averaged temperature, respectively, variables with the subscript 1 are complex, and Re[. . .] signifies the real part. The subscript 2,0 indicates a second-order time-independent quantity. The angular frequency of the oscillation is ω, t is time, and x is the coordinate along the axis of the regenerator, with $x=0$ at the ambient face and $x=x_w$ at the working-temperature face. If the perimeter of the regenerator is well insulated, the steady-state second-order energy flux, $\dot{H}_2$, is independent of both x and t.[14]

A simplified model of the acoustics in the regenerator, which we summarize here, has been used to obtain relationships between the various steady-state terms.[2] If the compliance, or void volume, of the regenerator is neglected, and the density oscillations are taken as isothermal, mass conservation and the temperature dependence of gas density require that the volumetric flow rate U_1 be given by

$$U_1(x)=U_1(0)\frac{T_m(x)}{T_a}, \tag{4}$$

where $T_a=T_m(0)$ is the mean temperature of the gas at the ambient end of the regenerator. Throughout the rest of this article, the subscripts "a" and "w" refer to a mean variable evaluated at the ambient and working-temperature faces of the regenerator, respectively. The volumetric flow rate into the ambient end, $U_1(0)$, is estimated[2] using a simplified model of the feedback network, yielding

$$U_1(0)=\frac{\omega^2 LC}{R_m}p_1(0), \tag{5}$$

where L and C are the inertance and compliance in the feedback loop, respectively. Typically, $\omega L<R_m$, and terms of order $(\omega L/R_m)^2$ have been dropped to simplify the calculation. Here, R_m is the flow resistance of the regenerator referenced to $U_1(0)$, i.e.,

$$R_m=\frac{\Delta p_{1,\mathrm{regen}}}{U_1(0)}=\frac{R_0}{x_w}\int_0^{x_w}\left(\frac{T_m(x)}{T_a}\right)^{1+b}dx, \tag{6}$$

where R_0 is the flow resistance of the regenerator when its entire length x_w is at T_a and $\Delta p_{1,\mathrm{regen}}$ is the change in p_1 across the regenerator.[2] The precise value of R_0 is not important in this calculation. The "1" in the exponent $(1+b)$ takes into account the T_m dependence of density and U_1 expressed in Eq. (4), and the "b" in the exponent takes into account the temperature-dependent viscosity of the gas in the regenerator, i.e., $\mu(T_m)=\mu(T_a)[T_m(x)/T_a]^b$. The expression for R_m is valid in the low-Reynolds-number limit for screen beds, parallel plates, and other typical regenerator geometries.[15] In most cases, $|\Delta p_{1,\mathrm{regen}}|\ll|p_1|$, and, therefore, $p_1(x)\approx p_1$. Without loss of generality, p_1 can be taken to be real. Equations (4) and (5) show that $U_1(x)$ is also real. Other symbols with subscript 1 will continue to represent complex variables.

The steady-state time-averaged second-order mass flux[16] is given by

$$\dot{M}_{2,0}=\mathrm{Re}[\rho_1\tilde{U}_1]/2+\rho_m U_{2,0}, \tag{7}$$

where ρ_m is the mean density of gas in the regenerator, ρ_1 is the first-order complex density oscillation amplitude, and the tilde denotes complex conjugation. Inside the regenerator, the pressure oscillations are nearly isothermal. Neglecting the small deviations from isothermal and using the fact that U_1 is real, the first term in $\dot{M}_{2,0}$ can be written[7] $(\rho_m/p_m)p_1U_1/2$, so that $\dot{M}_{2,0}=(\rho_m/p_m)p_1U_1/2+\rho_m U_{2,0}$. The second term in $\dot{M}_{2,0}$ depends on the second-order, steady volumetric flow rate $U_{2,0}$, which is driven by gradients in the second-order steady pressure $p_{2,0}$ and by source terms that involve time averages of two first-order quantities.[16,17] Waxler has shown[17,18] that the latter source terms are negligible compared to the gradients in $p_{2,0}$ in a parallel-plate regenerator, and we assume that this holds true for other regenerator-pore geometries. The second-order momentum equation is then identical to the first-order momentum equation with d/dt, $U_{2,0}$, and $p_{2,0}$ replacing $i\omega$, U_1, and p_1, respectively.

In a TASHE or FPTR, there can be several sources of $p_{2,0}$ including, but not limited to, pipe bends, diffusers, and "tees."[2,5] However, in a well-designed device, the major source of $p_{2,0}$ will be the jet pump used to control the streaming. In the steady state, $\dot{M}_{2,0}=0$ and $p_1(x)$ is assumed to be equal to $p_1(0)$, requiring $|U_{2,0}|$ to have the same spatial and $T_m(x)$ dependence as $|U_1|$. Therefore, the same definition of R_m also applies to $U_{2,0}$, i.e., $R_m=\Delta p_{2,0}/U_{2,0}(0)$. Using these two results, $\dot{M}_{2,0}$ at the ambient end of the regenerator can be written

$$\dot{M}_{2,0}(0)=\frac{\rho_a}{2p_m}p_1U_1(0)+\rho_a\frac{\Delta p_{2,0}}{R_m}. \tag{8}$$

Since $\partial\dot{M}_{2,0}/\partial x=0$, this also gives the value of $\dot{M}_{2,0}$ throughout the regenerator. Although $\dot{M}_{2,0}=0$, Eq. (7) is required later when the steady-state solution is perturbed. If the jet pump is located near the ambient end of the regenerator, the time-average pressure drop across the regenerator, $\Delta p_{2,0}$, will be given by[5]

$$\Delta p_{2,0}=-\frac{\rho_a[U_1(0)]^2}{8A_{jp}^2}(K_{\mathrm{out}}-K_{\mathrm{in}}), \tag{9}$$

where A_{jp} is the area of the small opening in the jet pump, and K_{in} and K_{out} are the minor loss coefficients for the two directions of flow through the jet pump. If the jet pump were located near the working-temperature heat exchanger instead of the ambient end, an additional multiplicative factor of T_w/T_a would appear in Eq. (9). As will be seen later [see, e.g., Eq. (36) and the surrounding discussion], any multiplicative scale factor in $\Delta p_{2,0}$ that is unaffected by the perturbation does not affect the final result of the calculation.

III. PERTURBED SOLUTION

Next, an exponentially growing or decaying perturbation is added to the equilibrium solution reviewed in the previous section, so that the complete solution is of the form

$$p(x,t)=p_m+p_1\cos\omega t+p_{2,0}(x)+\delta p_{2,0}(x)e^{\epsilon t}, \tag{10}$$

$$U(x,t)=U_1(x)\cos\omega t+U_{2,0}(x)+\{\delta U_1(x)\cos\omega t+\delta U_{2,0}(x)\}e^{\epsilon t}, \tag{11}$$

$$T(x,t)=T_m(x)+\mathrm{Re}[T_1(x)e^{i\omega t}]+\{\delta T_m(x)+\mathrm{Re}\,[\delta T_1(x)e^{i\omega t}]\}e^{\epsilon t}, \tag{12}$$

$$R(t)=R_m+\delta R\,e^{\epsilon t}, \tag{13}$$

$$\dot{H}(x,t)=\dot{H}_2+\delta\dot{H}(x)e^{\epsilon t}. \tag{14}$$

The perturbation includes both oscillating and nonoscillating terms. The oscillating terms are assumed to have the same frequency as the corresponding terms in the equilibrium solution and an amplitude that changes slowly, but exponentially, in time compared to the acoustic period, i.e., $|\epsilon|\ll\omega$. This two-time-scale approach allows the explicit separation of the slow change of the instability from the rapid acoustic oscillations. Nonoscillating terms also change exponentially in time with the same time constant as the amplitudes of the oscillating terms. Note that δU_1 can be taken to be real for the same reasons that U_1 is taken to be real.

We assume that T_a and T_w are fixed by good heat exchange with external fluids that have high heat capacity or latent heat or with high-heat-capacity heat exchangers. In either case, the thermal reservoirs in or near the heat exchangers are large enough to adjust to a varying heat load without a significant change in temperature. We also assume δp_1 and $\delta\omega$ are zero. For an FPTR, these last two conditions can be regarded as simple consequences of how the system is driven, e.g., by a linear motor at fixed frequency and whatever electric current is required to keep p_1 fixed. For a TASHE,[2] one can similarly assume that the complex load impedance is deliberately varied to keep ω and p_1 fixed.

The calculation begins with the energy flux $\dot{H}$ through the regenerator. In the steady state of a well-insulated regenerator, $d\dot{H}_2/dx=0$, meaning there is no build up of energy inside the regenerator.[14] If a small, time-dependent perturbation is present, this condition is relaxed and the energy equation takes the form

$$\frac{\partial}{\partial t}[(1-\phi)\rho_s c_s T_s A+\phi\rho_m c_v T_m A]=-\frac{\partial\dot{H}}{\partial x}, \tag{15}$$

where ρ_s, c_s, ϕ, A, and T_s are the regenerator solid density, heat capacity, porosity, cross-sectional area, and temperature, respectively, and c_v is the isochoric heat capacity of the gas. Essentially, this equation states that if there is a spatial variation in the energy flux through the regenerator, energy is accumulated or depleted temporally, resulting in a temperature change of the working gas and regenerator solid at that location. On the acoustic time scale, the temperature of the solid and gas are not necessarily equal. (In fact, oscillations of the gas temperature give rise to the major source of $\dot{H}_2$.[14]) However, on the slow time scale of the perturbation growth, the excellent thermal contact between the regenerator solid and the gas will enforce $\delta T_s\approx\delta T_m$. Also, in a typical regenerator, the heat capacity of the solid is much higher than the gas,[19] i.e., $\phi\rho_m c_v/(1-\phi)\rho_s c_s\ll 1$. Substituting the equilib-

rium solution plus perturbation into Eq. (15) and using the two approximations above yields, at linear order in the perturbation,

$$\epsilon(1-\phi)\rho_s c_s A\,\delta T_m = -\frac{d\,\delta\dot{H}}{dx}. \tag{16}$$

A typical solution of Eq. (16) would proceed as follows.[20] The steady-state solution plus perturbation would be substituted into the other governing equations, such as the momentum and continuity equations, and a system of differential equations relating $\delta\dot{H}(x)$ and $\delta T_m(x)$ would be obtained. Next, a set of eigenfunctions would be found that satisfies both Eq. (16) and any boundary conditions imposed at $x=0$ and x_w. These eigenfunctions would then be substituted back into Eq. (16) to determine the growth rate ϵ for each eigenfunction. Conditions under which one of the growth rates becomes positive would indicate an instability. Although this procedure would find all possible conditions for linear instability, carrying it out in this case would prove quite difficult. The resulting system of differential equations is not one with constant coefficients and is quite complicated. Finding a full set of eigenfunctions would be tedious, if not impossible, without numerical computation. Also, the difficulty of the calculation would obscure the physical effects that cause the instability.

To avoid this difficulty, a much simpler approach that yields analytical results is used, although it does not explore all possible instabilities. First, Eq. (16) is integrated from $x=0$ to $x=x_w$, eliminating the need to know the detailed form of $d\,\delta\dot{H}/dx$. Information about $\delta\dot{H}$ is only required at the ends of the regenerator. Next, the x dependence of the temperature perturbation is taken to be

$$\delta T_m = \sum_{n=1}^{n=\infty} \Theta_n \sin(n\pi x/x_w). \tag{17}$$

Implicit in Eq. (17) is our assumption that the heat exchangers at either end of the regenerator hold the faces at $x=0$ and $x=x_w$ at their steady-state values. It should be noted that the individual terms in Eq. (17) are not eigenfunctions of this problem, so there will be interaction between the various terms. However, this interaction will not be included in the following analysis, and only the $n=1$ term will be carried through. (Both of these simplifications are reasonable in light of the observed temperature distribution in the FPTR that demonstrated a streaming instability: The deviation from a linear temperature distribution resembled half of a sine wave.[11]) Substituting Eq. (17) into Eq. (16), setting $\Theta_1=\Theta$, and integrating from $x=0$ to $x=x_w$ yields

$$\frac{2x_w}{\pi}\rho_s c_s(1-\phi)A\epsilon\Theta - [\,\delta\dot{H}(0) - \delta\dot{H}(x_w)\,] = 0, \tag{18}$$

where any temperature dependence of $\rho_s c_s$ has been ignored.

Typically, regenerators are made from piles of stainless-steel mesh. However, to estimate which terms of $\delta\dot{H}$ are important, the analytic expression for $\dot{H}_2$ in a parallel-plate regenerator will be used. Except for axial conduction through the regenerator solid, the various terms are expected to have the same order-of-magnitude values in a screen-based regenerator. The steady-state energy flux consists of four terms[14]

$$\begin{aligned}\dot{H}_2(x) &= \frac{p_1 U_1}{2}\mathrm{Re}\left[1-\frac{f_\kappa-\tilde{f}_\nu}{(1+\sigma)(1-\tilde{f}_\nu)}\right] \\ &\quad + \frac{\rho_m c_p U_1^2\,\mathrm{Im}(f_\kappa+\sigma\tilde{f}_\nu)}{2\phi A\omega(1-\sigma^2)|1-f_\nu|^2}\frac{dT_m}{dx} \\ &\quad - [\,\phi Ak + (1-\phi)Ak_s]\frac{dT_m}{dx} + \dot{M}_{2,0}c_p T_m \\ &\equiv \dot{H}_b + \dot{H}_c + \dot{H}_k + \dot{H}_m,\end{aligned} \tag{19}$$

where each term has been given its own symbol for convenience, and the fact that U_1 is real has been used to simplify the expression. In the assumed steady-state solution, $\dot{M}_{2,0}=0$, and therefore $\dot{H}_m=0$. However, $\delta\dot{M}_{2,0}$ and its associated enthalpy flux, $\delta\dot{H}_m$, need not be zero. Since $\delta T_m(0)=\delta T_m(x_w)=0$, the gas properties and f functions at $x=0$ and $x=x_w$ do not change, and $\delta\dot{H}(0)$ can be written

$$\begin{aligned}\delta\dot{H}(0) &= \dot{H}_b(0)\frac{\delta U_1(0)}{U_1(0)} + \dot{H}_c(0)\left[2\frac{\delta U_1(0)}{U_1(0)}\right. \\ &\quad \left. + \frac{1}{\nabla T_m|_{x=0}}\left.\frac{d\,\delta T_m}{dx}\right|_0\right] + \frac{\dot{H}_k(0)}{\nabla T_m|_{x=0}}\left.\frac{d\,\delta T_m}{dx}\right|_0 \\ &\quad + \delta\dot{M}_{2,0}(0)c_p T_a \\ &\equiv \delta\dot{H}_b(0) + \delta\dot{H}_c(0) + \delta\dot{H}_k(0) + \delta\dot{H}_m(0),\end{aligned} \tag{20}$$

and similarly for $\delta\dot{H}(x_w)$.

As long as $2\,\delta U_1(0)/U_1(0)$ and $(d\,\delta T_m/dx|_{x=0})/\nabla T_m|_{x=0}$ do not cancel or sum to a value much smaller than either individual term, $\delta H_b(0)$, $\delta\dot{H}_c(0)$, and $\delta\dot{H}_k(0)$ can be compared simply by comparing the magnitudes of $\dot{H}_b$, $\dot{H}_c$, and $\dot{H}_k$. For $T_w<T_a$, both terms have the same sign. However, for $T_w>T_a$, they have opposite signs. In this case, Eqs. (5), (17), (29), and (37) can be used to show that the sum of theses two terms is never small compared to the either of individual terms themselves.

To compare $\dot{H}_c$ and $\dot{H}_k$, $\dot{H}_c$ is expanded in the limit $(r_h/\delta_\kappa)^2 \approx (r_h/\delta_\nu)^2 \ll 1$, where δ_κ and δ_ν are the thermal and viscous penetration depths and r_h is hydraulic radius.[21] This expansion is equally valid at low Reynolds numbers in screen-based regenerators where r_h is a rough measure of the average spacing between screen wires. In typical regenerators $(r_h/\delta_\kappa)^2 \lesssim 0.1$. After a small amount of algebra, the result is found to be

$$\dot{H}_c \approx -\frac{\rho_m c_p U_1^2}{2\phi A\omega}\left(\frac{r_h}{\delta_\kappa}\right)^2\frac{dT_m}{dx}. \tag{21}$$

Taking the ratio of $\dot{H}_k(0)$ with $\dot{H}_c(0)$ yields

$$\frac{\dot{H}_k(0)}{\dot{H}_c(0)} \approx 2\left(\frac{1-\phi}{\phi}\right)\left[\frac{\delta_{\kappa,a}^2}{\xi_1(0)r_h}\right]^2\frac{k_s}{k}, \tag{22}$$

where $\xi_1(0)$ is the acoustic displacement amplitude and all terms are evaluated at the ambient end of the regenerator. Substituting typical numbers $k_s/k \approx 100$, $(\delta_\kappa/r_h)^2 \approx 10$, $\delta_\kappa/\xi_1(0) \approx 0.01$, and $\phi \approx 0.7$ yields $\dot{H}_k/\dot{H}_c \approx 0.1$. In a screen-bed regenerator, this ratio is expected to be even smaller due to the poor thermal contact between screen layers.[22] Therefore, $\delta\dot{H}_k(0) \ll \delta\dot{H}_c(0)$ and can be neglected in the right-hand side of Eq. (20).

To compare $\delta\dot{H}_b(0)$ with $\delta\dot{H}_c(0)$, $\dot{H}_b$ is expanded in the limit $(r_h/\delta_\kappa)^2 \to 0$,

$$\dot{H}_b \approx \left(\frac{r_h}{\delta_\kappa}\right)^4 \frac{p_1 U_1}{3}. \tag{23}$$

Taking the ratio of $\dot{H}_b$ with $\dot{H}_c$ at the ambient end of the regenerator yields

$$\left|\frac{\dot{H}_b(0)}{\dot{H}_c(0)}\right| \approx 4\phi(\gamma-1)\Gamma\left(\frac{r_h}{\delta_{\kappa,a}}\right)^2 \frac{x_w}{\lambda_a}\frac{T_a}{\Delta T_m}. \tag{24}$$

Here, γ is the ratio of specific heats, λ is the acoustic wavelength, ΔT_m is the total temperature difference across the regenerator, $\Gamma = (p_1/U_1)/(\rho_a c_a/A)$, and ρ_a and c_a are the gas density and speed of sound at the ambient end. Substituting typical numbers $\phi \approx 0.7$, $\gamma - 1 \approx 0.67$, $(r_h/\delta_\kappa)^2 \approx 0.1$, $x_w/\lambda \approx 0.005$, $T_a/\Delta T_m \approx 1$, and $\Gamma \approx 10$ yields $|\dot{H}_b(0)/\dot{H}_c(0)| \approx 0.01$. Therefore, $\delta\dot{H}_b(0) \ll \delta\dot{H}_c(0)$ and can be neglected in the right-hand side of Eq. (20).

Finally, $\delta\dot{H}_c(0)$ is compared with $\delta\dot{H}_m(0)$. More specifically, $\dot{H}_c(0)\delta U_1(0)/U_1(0)$ is compared with $\delta\dot{M}_{2,0}(0)c_p T_a$. To estimate $\delta\dot{M}_{2,0}(0)$, only one component is used, namely $\mathrm{Re}[\rho_1(0)\delta\tilde{U}_1(0)]/2$. Assuming an isothermal density oscillation, the ratio of the two terms is

$$\frac{\delta\dot{M}_{2,0}(0)c_p T_a}{\dot{H}_c(0)\delta U_1(0)/U_1(0)} \approx 10\frac{\gamma\phi}{\sigma}\frac{x_w}{\lambda}\left(\frac{\delta_\nu}{r_h}\right)^2 \frac{T_a}{\Delta T_m}\Gamma \approx 10, \tag{25}$$

where the numerical value results from substituting typical numbers given previously. The approximation $\delta\dot{M}_{2,0}(0) \approx \mathrm{Re}[\rho_1(0)\delta\tilde{U}_1(0)]/2$ most likely overestimates $\delta\dot{M}_{2,0}(0)$.

Now that $\delta\dot{H}_c$ and $\delta\dot{H}_m$ have been established as the dominant terms in $\delta\dot{H}$, it remains to calculate these terms at either end of the regenerator. At $x=0$ and $x=x_w$, $\delta\dot{H}_c$ is given by

$$\delta\dot{H}_c(0) = \dot{H}_c(0)\left[2\frac{\delta U_1(0)}{U_1(0)} + \frac{1}{\nabla T_m|_0}\left.\frac{d\,\delta T_m}{dx}\right|_0\right], \tag{26}$$

$$\delta\dot{H}_c(x_w) = \dot{H}_c(x_w)\left[2\frac{\delta U_1(x_w)}{U_1(x_w)} + \frac{1}{\nabla T_m|_{x_w}}\left.\frac{d\,\delta T_m}{dx}\right|_{x_w}\right]. \tag{27}$$

In the steady state, $\dot{H}_c \gg \dot{H}_b$, $\dot{H}_c \gg \dot{H}_k$ and $\dot{H}_m = 0$. Therefore, $\dot{H}_c(0) \approx \dot{H}_c(x_w) \equiv \dot{H}_c$. Also, $\delta U_1(0)/U_1(0) = \delta U_1(x_w)/U_1(x_w)$. The difference in $\delta\dot{H}_c$ across the regenerator, needed for Eq. (18), is then

$$\delta\dot{H}_c(0) - \delta\dot{H}_c(x_w) = \dot{H}_c\left[\frac{1}{\nabla T_m|_0}\left.\frac{d\,\delta T_m}{dx}\right|_0 - \frac{1}{\nabla T_m|_{x_w}}\left.\frac{d\,\delta T_m}{dx}\right|_{x_w}\right]. \tag{28}$$

To be consistent with $\dot{H}_c(0) = \dot{H}_c(x_w)$, Eq. (21) shows that the mean temperature gradients at the two ends are related by

$$\nabla T_m|_{x_w} = \tau^b \nabla T_m|_0, \tag{29}$$

where $\tau = T_w/T_a$. Using this in Eq. (28), the difference in $\delta\dot{H}_c$ is found to be

$$\delta\dot{H}_c(0) - \delta\dot{H}_c(x_w) = \frac{\dot{H}_c}{\nabla T_m|_0}\frac{\pi}{x_w}(1+\tau^{-b})\Theta. \tag{30}$$

For both an FPTR and a TASHE, the term $\dot{H}_c/\nabla T_m|_0$ is negative. Therefore, if the perturbation of the regenerator temperature is positive ($\Theta > 0$), the effect of $\delta\dot{H}_c$ is to remove energy and cool the regenerator. This effect, present in any Stirling-type engine or refrigerator whether or not a toroidal topology allows streaming, attempts to reduce any perturbation from the steady state, suppressing a possible instability.

Next, the effect of $\delta\dot{H}_m$ is determined. The difference in $\delta\dot{H}_m$ at the two regenerator faces is given by

$$\delta\dot{H}_m(0) - \delta\dot{H}_m(x_w) = \delta\dot{M}_{2,0}(0)c_p T_a - \delta\dot{M}_{2,0}(x_w)c_p T_w. \tag{31}$$

To proceed further, a relationship between $\delta\dot{M}_{2,0}(0)$ and $\delta\dot{M}_{2,0}(x_w)$ must be determined. In the steady-state solution, the time-averaged continuity equation[16] reveals that since $\partial\rho_m/\partial t = 0$, $\dot{M}_{2,0}(0) = \dot{M}_{2,0}(x_w)$. For $\delta\dot{M}_{2,0}$, it is not clear if this relationship still holds. Let $\alpha = \delta\dot{M}_{2,0}(0) - \delta\dot{M}_{2,0}(x_w)$. If $\alpha \neq 0$, then the density of the gas in the regenerator must be changing, i.e., $\alpha \approx \phi A x_w \epsilon \delta\rho_m$. This change in density is driven by either a change in mean temperature or a change in mean pressure. Considering a change in mean temperature,

$$\alpha \approx \phi A x_w \rho_m \frac{\epsilon\,\delta T_m}{T_m}. \tag{32}$$

However, for δT_m to change in the first place, there must be a $\delta\dot{M}_{2,0}$ given by

$$(1-\phi)\rho_s c_s A x_w \epsilon\,\delta T_m \approx \delta\dot{M}_{2,0}c_p(T_w - T_a). \tag{33}$$

Combining the last two equations yields

$$\frac{\alpha}{\delta\dot{M}_{2,0}} \approx \frac{\phi\rho_m c_p}{(1-\phi)\rho_s c_s}\frac{T_w - T_a}{T_m} \ll 1, \tag{34}$$

showing that, to a good approximation, $\delta\dot{M}_{2,0}(0) = \delta\dot{M}_{2,0}(x_w)$. Therefore, the difference in $\delta\dot{H}_m$ at the two ends of the regenerator can be written

$$\delta\dot{H}_m(0) - \delta\dot{H}_m(x_w) = \delta\dot{M}_{2,0}(0)c_p T_a(1-\tau). \tag{35}$$

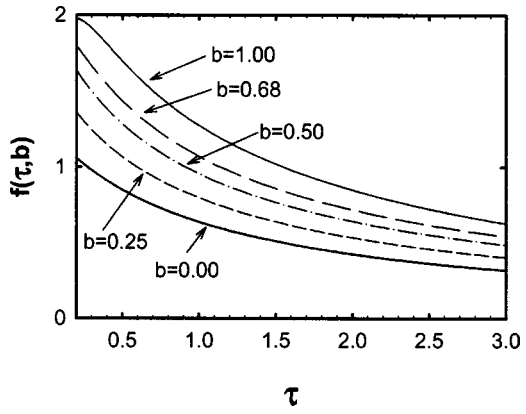


FIG. 2. A plot of $f(\tau,b)$ for various values of b. For helium, $b\approx 0.68$. The integrals in Eq. (37) are performed numerically.

Perturbing Eq. (7) and using $\dot{M}_{2,0}=0$ in the steady state, $\delta U_1(0)/U_1(0)=-\delta R/R_m$ from Eq. (5), and $\delta\Delta p_{2,0}/\Delta p_{2,0}=-2\,\delta R/R_m$ from Eq. (9) yields

$$\delta\dot{M}_{2,0}(0)=-\frac{\rho_a}{p_m}\frac{p_1 U_1(0)}{2}\frac{\delta\Delta p_{2,0}}{\Delta p_{2,0}}=\frac{2\rho_a}{p_m}\dot{E}_2(0)\frac{\delta R}{R_m}, \tag{36}$$

where $\dot{E}_2(0)$ is the acoustic power flowing into the ambient end of the regenerator. $\delta R/R_m$ is calculated by perturbing Eq. (6):

$$\frac{\delta R}{R_m}=\frac{(1+b)\int_0^{x_w} T_m^b(x)\sin(\pi x/x_w)dx}{\int_0^{x_w} T_m^{1+b}(x)dx}\Theta=\frac{\Theta}{T_a}f(\tau,b), \tag{37}$$

where, to be consistent with Eq. (29), $T_m(x)$ is approximated by

$$T_m(x)=T_a\left[1+(\tau-1)\frac{x}{x_w}-\frac{\tau-1}{\pi}\frac{\tau^b-1}{\tau^b+1}\sin\left(\frac{\pi x}{x_w}\right)\right], \tag{38}$$

and the variable of integration is changed to $d(x/x_w)$. A plot of $f(\tau,b)$ for various values of b is shown in Fig. 2. Finally, combining Eqs. (35), (36), and (37) yields

$$\delta\dot{H}_m(0)-\delta\dot{H}_m(x_w)=\frac{2\gamma}{\gamma-1}(1-\tau)f(\tau,b)\dot{E}_2(0)\frac{\Theta}{T_a}. \tag{39}$$

From Eq. (39) it is clear what drives the instability. For $\Theta>0$, the sign of the right-hand side of this expression is determined by $(1-\tau)$. All other factors are positive. For a TASHE, $(1-\tau)$ is negative and the perturbed time-average mass flux removes energy from the regenerator, reducing the original perturbation. For an FPTR, $(1-\tau)$ is positive and energy is dumped into the regenerator, amplifying the perturbation. Equation (30) already showed that no matter the value of τ, $\delta\dot{H}_c$ reduces the original perturbation. For $\tau>1$, i.e., a TASHE, a jet pump can reliably be used to suppress streaming. For $\tau<1$, i.e., an FPTR, there exists a possibility of an instability with $\delta\dot{H}_c$ competing with $\delta\dot{H}_m$.

Now that the term that drives the instability is identified, the cause can be investigated. In either an FPTR or a TASHE, if Θ is positive, the average temperature of the regenerator increases slightly and its resistance grows, reducing $U_1(0)$. The effect on the two terms in Eq. (7) is slightly different. The smaller $U_1(0)$ results in a lower $\mathrm{Re}[\rho_1(0)\tilde{U}_1(0)]/2$, but $|U_{2,0}|$ decreases even more because it varies as $U_1^2(0)$ through its dependence on $\Delta p_{2,0}$. Therefore, when the regenerator warms slightly, $\dot{M}_{2,0}$ increases slightly, i.e., $\delta\dot{M}_{2,0}(0)>0$. In a TASHE, this blows ambient gas into the regenerator, cooling it down and suppressing any instability. In an FPTR, the ambient gas warms the regenerator further, creating an instability.

Combining Eqs. (18), (30), and (39) gives an equation for the growth rate of the perturbation

$$\frac{2}{\pi}\rho_s c_s T_a(1-\phi)Ax_w\epsilon=\frac{\pi(T_a/x_w)}{\nabla T_m|_0}(1+\tau^{-b})\dot{H}_c+\frac{2\gamma}{\gamma-1}(1-\tau)f(\tau,b)\dot{E}_2(0). \tag{40}$$

If the right-hand side of this equation is positive, the perturbation will grow exponentially and an instability results with a large streaming mass flux around the lumped-element loop. If it is negative, the perturbation decays and the jet pump controls the streaming in a stable manner. As already discussed, both terms on the right-hand side are negative for a TASHE. Therefore, a jet pump can be used in a TASHE with confidence that it will always control the streaming. In an FPTR, the first term is negative and the second is positive. Therefore, they must be compared to see when an instability results.

To make this comparison more transparent, $\dot{E}_2(0)$ is expressed in terms of the gross cooling power of the FPTR, i.e., $\dot{Q}_{c,\mathrm{gross}}\approx\tau\dot{E}_2(0)$.[6] The right-hand side of Eq. (40) is positive when

$$\frac{\dot{H}_c}{\dot{Q}_{c,\mathrm{gross}}}<\frac{2}{\pi}\frac{\gamma}{\gamma-1}\frac{(\tau-1)f(\tau,b)}{(1+\tau^{-b})\tau}\frac{\nabla T_m|_0}{(T_a/x_w)}. \tag{41}$$

Because $\nabla T_m|_{x_w}=\tau^b\nabla T_m|_0$, the term $\nabla T_m|_0/(T_a/x_w)$ is given by $2(\tau-1)/(1+\tau^b)$. Substituting this result into Eq. (41), the final result is

$$\frac{\dot{H}_c}{\dot{Q}_{c,\mathrm{gross}}}<\frac{4}{\pi}\frac{\gamma}{\gamma-1}\frac{(\tau-1)^2 f(\tau,b)}{(1+\tau^b)(1+\tau^{-b})\tau}. \tag{42}$$

The right-hand side of Eq. (42) is computed as a function of τ for $\gamma=5/3$ and $b=0$, $b=0.68$ (typical for helium), and $b=0.85$ (typical for argon). The results are shown in Fig. 3. For FPTRs operating below and to the left of the lines, a jet pump *cannot* be used to control streaming in stable fashion. For FPTRs operating above and to the right of the lines, a jet pump will suppress streaming in a stable fashion. The open squares in Fig. 3 are two different operating points from a $\approx$2-kW FPTR intended to liquefy natural gas.[11] The jet pump in this FPTR never was able to reliably control the streaming. Since this FPTR used helium as its working gas, the squares should be compared to the $b=0.68$ line. The circles are from an earlier benchtop FPTR.[5] The filled circles

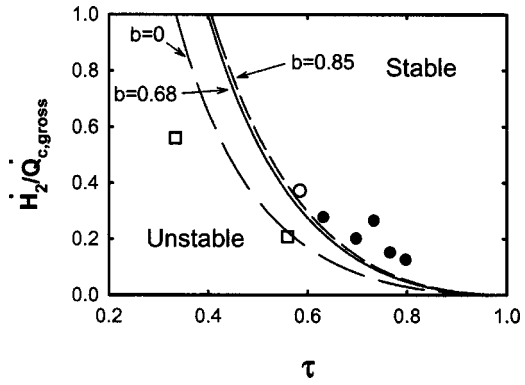


FIG. 3. Stability curve. The lines are the threshold of instability when the left-hand side of Eq. (42) equals the right-hand side for $b=0$, $b=0.68$, and $b=0.85$. As shown by Eqs. (22) and (24), $\dot{H}_k$ and $\dot{H}_b$ are much smaller than $\dot{H}_c$ so that $\dot{H}_2 \approx \dot{H}_c$. Below and left of these lines, a jet pump will not suppress streaming in a stable manner. The open squares left of the line are data from an FPTR that demonstrated a streaming instability (Ref. 11) and should be compared with the $b=0.68$ line. The circles are from an earlier benchtop FPTR (Ref. 5). The filled circles are operating points where the jet pump controlled the streaming stably. The open circle is an operating point where the streaming control was unstable. The circles should be compared to the $b=0.85$ line. In both cases, $\dot{H}_2$ and $\dot{Q}_{c,\mathrm{gross}}$ are not directly measured. They are calculated using DeltaE (Ref. 24).

are operating points where the streaming control was stable. The open circle is an operating point with unstable streaming control. This FPTR used argon as a working gas, so the circles should be compared with the $b=0.85$ line.

To be consistent with the steady-state condition $\dot{H}_c(0) = \dot{H}_c(x_w)$ and the neglect of regenerator compliance in Eq. (5), the approximate mean temperature given in Eq. (38) was used throughout the calculation. However, in actual hardware where the compliance of the regenerator may be important or the phase of $U_1(0)$ relative to p_1 may be significantly different from zero, $T_m(x)$ may deviate from the temperature profile used in the calculation. The calculation has been redone for a linear temperature profile, i.e., a constant $\nabla T_m(x)$. There are differences with the results shown in Fig. 3, but they are small enough not to be noticeable on the plot. The result that a TASHE is always stable still holds for a linear temperature profile as well.

IV. CONCLUSION

Using a simplified model of the acoustics in the regenerator of a TASHE and FPTR, the stability of jet pump control of streaming has been investigated. A stability criterion has been derived and found to be in agreement with the meager data available to date. The stability criterion shows that jet pump control of streaming is stable for all TASHEs. It also provides a threshold temperature ratio below which streaming control in a FPTR is unstable. The mathematics is based on analyzing how the temperature in the center of the regenerator responds to changes in streaming flows through the regenerator, which are themselves controlled by the temperature in the center of the regenerator. The analysis shows that two effects dominate.

First, ordinary second-order energy flow through the regenerator, whose largest term is proportional to $U_1^2 dT_m/dx$, always exerts a stabilizing influence. If the temperature of the center of the regenerator of a TASHE decreases a small amount, enthalpy flow from the hot end to the center increases as the average temperature gradient in the hot half steepens, and enthalpy flow from the center to the ambient end decreases as the average temperature in the cooler half becomes more shallow; both of these changes in enthalpy flow tend to raise the temperature in the center, canceling the original, assumed decrease in temperature. Similar arguments for an assumed small increase in the center temperature, and for an FPTR instead of a TASHE, also lead to cancellation, and, hence, stability.

Second, the temperature dependences of the viscosity and density of the gas in the regenerator cause a change in streaming that affects that very temperature. In an FPTR, this effect is destabilizing. If the temperature of the center of the regenerator of an FPTR decreases a small amount, the viscosity decreases and the density increases; both of these changes reduce R, leading to increases in both U_1 and $U_{2,0}$. If $\Delta p_{2,0}$ exerted by the jet pump were to remain constant, the fractional changes in U_1 and $U_{2,0}$ would be equal to the fractional change in R, and the streaming—a balance between U_1 and $U_{2,0}$—would change little. However, $\Delta p_{2,0}$ does not remain constant; it increases, thereby changing the streaming in a direction that carries cold gas into the regenerator, amplifying the original temperature decrease. A similar argument for a TASHE leads instead to stability.

Operation of a cryogenic FPTR with deliberately large $\dot{H}_2$ to enforce stability is very undesirable, because nonzero $\dot{H}_2$ consumes cooling power, reducing efficiency. However, the present analysis hints at ways that the stability curve might be shifted slightly. Three examples will be mentioned. First, the analysis assumed that K_{out}, K_{in}, and A_{jp} in Eq. (9) are independent of amplitude. If one or more of these coefficients depended on U_1, either via hydrodynamics or elastic motion of the jet pump walls, a region of enhanced stability could be created. Second, Eq. (42) shows that reduced γ or increased b improves stability. Third, the analysis assumed that R, f_κ, and f_ν are independent of velocity, but the more complicated, velocity-dependent flow resistance and heat transfer coefficient in screen beds may provide an opportunity for improved stability.

ACKNOWLEDGMENTS

This work was supported by the OBES/DMS in the U. S. Department of Energy's Office of Science.

[1] S. Backhaus and G. W. Swift, "A thermoacoustic-Stirling heat engine," Nature (London) **399**, 335–338 (1999).
[2] S. Backhaus and G. W. Swift, "A thermoacoustic-Stirling heat engine: Detailed study," J. Acoust. Soc. Am. **107**, 3148–3166 (2000).
[3] T. Yazaki, A. Iwata, T. Maekawa, and A. Tominaga, "Traveling wave thermoacoustic engine in a looped tube," Phys. Rev. Lett. **81**, 3128–3131 (1998).
[4] C. M. de Blok and N. A. H. J. van Rijt, Thermo-acoustic system, U. S. Patent No. 6,314,740, 13 Nov. 2001.
[5] G. W. Swift, D. L. Gardner, and S. Backhaus, "Acoustic recovery of lost power in pulse tube refrigerators," J. Acoust. Soc. Am. **105**, 711–724 (1999).
[6] P. Kittel, "Ideal orifice pulse tube refrigerator performance," Cryogenics **32**, 843–844 (1992).

[7] D. Gedeon, ''dc gas flows in Stirling and pulse-tube cryocoolers,'' in *Cryocoolers 9*, edited by R. G. Ross (Plenum, New York, 1997), pp. 385–392.
[8] S. Backhaus and G. W. Swift, ''Fabrication and use of parallel plate regenerators in thermoacoustic engines,'' in *Proceedings of the 36th Intersociety Energy Conversion Engineering Conference* (American Society of Mechanical Engineers, New York, 2001), pp. 453–458.
[9] S. Backhaus, E. Tward, and M. Petach, ''Thermoacoustic power systems for space applications,'' in *Proceedings of the Space Technology and Applications International Forum 2002*, edited by M. El-Genk (Springer, New York, 2002), pp. 939–944.
[10] Unpublished data from 2000 collaboration between Chart, Inc. and Los Alamos National Laboratory.
[11] Unpublished data from 1999 collaboration between Chart, Inc. and Los Alamos National Laboratory.
[12] J. R. Olson, V. Kotsubo, P. J. Champagne, and T. C. Nast, ''Performance of a Two-State Pulse Tube Cryocooler for Space Applications,'' in *Cryocoolers 10*, edited by R. G. Ross (Plenum, New York, 1999), pp. 163–170.
[13] V. Kotsubo, P. Huang, and T. C. Nast, ''Observation of dc flows in a double inlet pulse tube,'' in *Cryocoolers 10*, edited by R. G. Ross (Plenum, New York, 1999), pp. 299–305.
[14] G. W. Swift, *Thermoacoustics: A Unifying Perspective for some Engines and Refrigerators* (Acoustical Society of America, NY, 2002).
[15] In regenerators, typical Reynolds numbers are on the order of 10 to 100. For parallel plates and circular or rectangular pores, the low-Reynolds number limit applies up to Reynolds numbers of ≈ 1000 or higher (Ref. 23) where a relatively sharp transition from laminar to turbulent flow is observed. However, the transition is not so sharp in screen beds. In this case, the flow can be characterized by a ''laminar'' flow resistance, i.e., one that does not depend on Reynolds number, and a turbulent flow resistance that increases linearly with Reynolds number (Ref. 19). For screen bed porosities in the range 0.65 to 0.75, the laminar and turbulent contributions are equal for Reynolds numbers of 90 to 130. This is the upper range of Reynolds numbers in regenerators. The data presented in this article are taken at Reynolds numbers of ≈ 20 (Ref. 5) and ≈ 45 (Ref. 11).
[16] W. L. M. Nyborg, ''Acoustic streaming,'' in *Physical Acoustics*, edited by W. P. Mason (Academic, New York, 1965), Vol. IIB, p. 265.
[17] R. Waxler, ''Stationary velocity and pressure gradients in a thermoacoustic stack,'' J. Acoust. Soc. Am. **109**, 2739–2750 (2001).
[18] Waxler (Ref. 17) shows that $dp_{2,0}/dx$ dominates other contributions to $U_{2,0}$ by a factor of order $(\delta_\kappa/r_h)^2$, where δ_κ is the thermal penetration depth and $2r_h$ is the gap in the parallel-plate regenerator. This factor is always large for regenerators that are at least reasonably effective. The neglect of the temperature dependence of viscosity is not always acceptable in streaming calculations [see, e.g., J. R. Olson and G. W. Swift, ''Acoustic streaming in pulse tube refrigerators: Tapered pulse tubes,'' Cryogenics **37**, 769–776 (1997)]. Incorporating this effect into Waxler's analysis adds a term $\langle(\partial/\partial y)\mu_1 \partial v_{1x}/\partial y\rangle$ to his Eq. (21), in his notation. This term also turns out to be only of order $(r_h/\delta_\kappa)^2$.
[19] G. W. Swift and W. C. Ward, ''Simple harmonic analysis of regenerators,'' J. Thermophys. Heat Transfer **10**, 652–662 (1996).
[20] S. Chandrasekhar, *Hydrodynamic and Hydromagnetic Stability* (Dover, New York, 1961).
[21] The hydraulic radius is the ratio of the gas volume to the gas–solid contact area. For example, for parallel plates r_h is half of the distance between the plates.
[22] M. A. Lewis, T. Kuriyama, F. Kuriyama, and R. Radebaugh, ''Measurement of heat conduction through stacked screens,'' Adv. Cryog. Eng. **43**, 1611–1618 (1998).
[23] W. M. Kays and A. L. London, *Compact Heat Exchangers* (McGraw-Hill, New York, 1964).
[24] W. C. Ward and G. W. Swift, ''Design environment for low amplitude thermoacoustic engines (DeltaE),'' J. Acoust. Soc. Am. **95**, 3671–3672 (1994). Software and user's guide available either from the Los Alamos thermoacoustics website at www.lanl.gov/thermoacoustics/ or from the Energy Science and Technology Software Center, US Department of Energy, Oak Ridge, Tennessee.

Study of critical behavior in concrete during curing by application of dynamic linear and nonlinear means

Jean-Christoph Lacouture[a)]
Laboratoire Environnement et Développement, Université D. Diderot, Tour 33-43, Case courrier 7087, 2 Place Jussieu, 75251 Paris Cedex, France

Paul A. Johnson[b)]
Geophysics, MS D443, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Frederic Cohen-Tenoudji[c)]
Laboratoire Environnement et Développement, Université D. Diderot, Tour 33-43, Case courrier 7087, 2 Place Jussieu, 75251 Paris Cedex, France

(Received 7 March 2002; revised 28 October 2002; accepted 16 December 2002)

The monitoring of both linear and nonlinear elastic properties of a high performance concrete during curing is presented by application of compressional and shear waves. To follow the linear elastic behavior, both compressional and shear waves are used in wide band pulse echo mode. Through the value of the complex reflection coefficient between the cell material (Lucite) and the concrete within the cell, the elastic moduli are calculated. Simultaneously, the transmission of a continuous compressional sine wave at progressively increasing drive levels permits us to calculate the nonlinear properties by extracting the harmonics amplitudes of the signal. Information regarding the chemical evolution of the concrete based upon the reaction of hydration of cement is obtained by monitoring the temperature inside the sample. These different types of measurements are linked together to interpret the critical behavior. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1543927]

PACS numbers: 43.25.Zx, 43.25.Ba, 43.25.Ts, 43.35.Zc [MFH]

I. INTRODUCTION

Concrete, especially Ultra High Performance Concrete (UHPC), is well known to be a complex, multiscale material.[1] The physical properties of concrete evolve significantly during curing, experiencing a phase change, a critical behavior, from the liquid to solid state. The hardening of concrete caused by the reaction of hydration of the cement is a well-studied process[2] and there are many papers dealing with the linear elastic properties of concrete during curing.[3,4] Furthermore, it is known that hardened concrete shows nonlinear properties[5] related in some manner to the presence of microcracking. We are not aware of any studies that monitor both the linear and dynamic nonlinear elastic parameters through a phase change. Nonlinear techniques have the potential to help us unravel the complex physical property changes that take place during a phase change. Concrete is a complex and interesting candidate for study because the phase change takes place over several hours and therefore one can probe the material many times during the change.

This paper begins with a description of the experimental configuration, followed by a description of experimental results, discussion, and conclusions.

[a)]Electronic mail: lacoutur@ccr.jussieu.fr
[b)]Electronic mail: paj@lanl.gov
[c)]Electronic mail: tenoudji@ccr.jussieu.fr

II. EXPERIMENTAL PROCEDURE

A cubic-shaped volume $16\times16\times16$ cm^3 constructed of wooden walls and a base of Lucite is used for the experiment (Fig. 1). High Performance Reactive Powder Concrete composed of Portland cement, thin sand, silica fume, water, and superplasticizer (Tables I and II), is poured into the box, and the monitoring commences immediately. Two wide band piezoelectric transducers of 500 kHz central frequency, are attached to the Lucite base of the experiment cell. One transducer generates compressional and the other shear waves. Both transducers are operated in pulse echo mode. From the measured reflection coefficients and the density, we can calculate the elastic moduli of the concrete (see below) and follow their evolution with time. In order to obtain the nonlinear response, we transmit pure-tone compressional waves at 8 kHz across the sample at 20 successively increasing amplitudes. The piezoceramic (PZT) emitter and receiver are driven at 8 kHz, far from their natural resonant frequency (100 kHz and 200 kHz, respectively) and placed in direct contact with the concrete through holes in the cell walls. We monitor the amplitudes of harmonics in the transmitted signal as a function of the amplitude of the fundamental. In order to avoid interfering effects, the linear and nonlinear measurements are performed at successively different times over the duration of the experiment. That is, the reflection coefficient measurement is conducted, and following this, the nonlinear transmission data are collected, and so on at 5 min intervals. Data are stored and analyzed on the computer. The

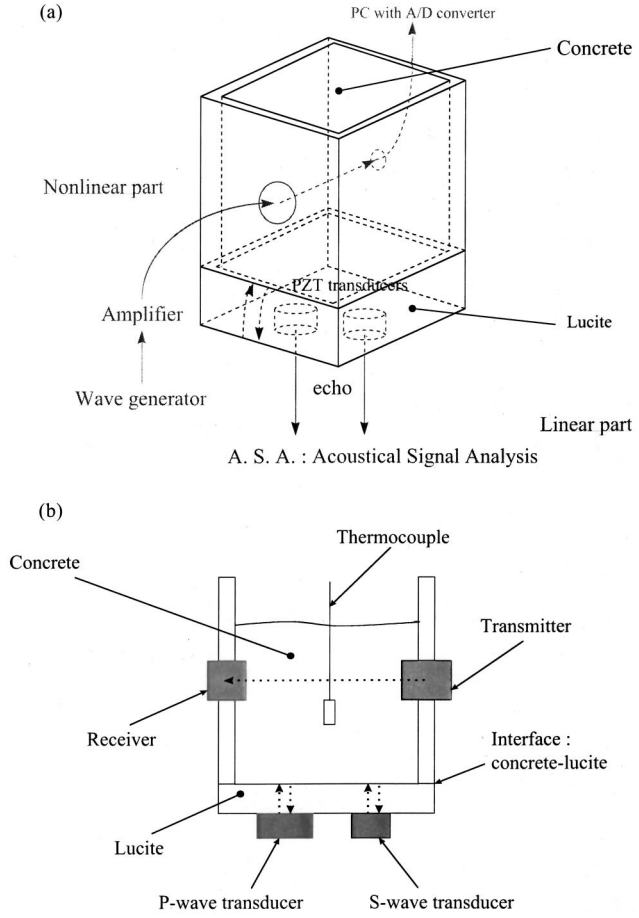


FIG. 1. (a) Bird's-eye view of experimental container showing location of transducers. (b) side view of experimental container. The nonlinear signals are obtained from the waves traversing the box at its midpoint. The wavespeeds are obtained from the reflection coefficients of the *P*- and *S*-waves between the Lucite base and the concrete shown at the container base.

temperature is monitored using a thermocouple embedded in the sample and the ambient temperature is measured with a separate thermocouple.

III. THEORETICAL CONSIDERATIONS

A. Linear elastic moduli

The Young's modulus E and Poisson's ratio υ are calculated from the reflection coefficient of longitudinal and shear waves. In an homogeneous medium of density ρ, the wave equation of the displacement U in one dimension x is given by

TABLE I. Cement paste composition.

Component	Density	Dry mass ratio (normalized to cement)	Granulometry (μm)
HTS cement	3.17	1	20
BE31 sand	2.65	1.1	310
MST fume silica	2.27	0.25	0.2
Mapei X404 superplasticizer	1.1	0.011	
Water	1	0.213	

TABLE II. Chemical composition of HTS cement (%).

	HTs cement
$C_2S(3CaOH,SiO_2)$	62–64
$C_2S(2CaOH,SiO_2)$	18–21
$C_2A(3CaOH,Al_2O_3)$	3.8–4.4
$C_4AF(4CaOH,Al_2O_3,Fe_2O_3)$	6.5–7.4

$$\rho \frac{\partial^2 U}{\partial t^2} = M \frac{\partial^2 U}{\partial x^2}, \tag{1}$$

where $M=\lambda+2\mu$ for longitudinal waves and $M=\mu$ for shear waves, with λ and μ being the Lamé coefficients. In the case of a viscoelastic material M is complex. For harmonic waves, Eq. (1) becomes

$$-\rho\omega^2 U = -Mk^2 U, \tag{2}$$

where k is the wave number,

$$\frac{\rho\omega^2}{M} = k^2. \tag{3}$$

Writing $k=k'+ik''$, the phase velocity is given by

$$c = \frac{\omega}{k'}. \tag{4}$$

The stress amplitude reflection coefficient is given by

$$r = \frac{z_2 - z_1}{z_2 + z_1}, \tag{5}$$

where z_1 is the acoustic impedance of the first medium, here the Lucite, and z_2 is the acoustic impedance of the second medium, the concrete.

The impedance is given by

$$z = \sqrt{\rho M}. \tag{6}$$

In the frequency range 100 kHz–600 kHz, the acoustic impedance z_1 of the Lucite remains mostly real. The values are $z_1=2\times10^6$ Rayl for longitudinal waves and $z_1=10^6$ Rayl for shear waves. It is observed that the impedance of the concrete z_2 happens to be highly complex at certain stages of the cure.

The complex reflection coefficient r is obtained in the frequency domain by applying a Fourier transformation (at 300 kHz) of the reflected pulse at the interface concrete-Lucite, normalized in modulus and phase by the amplitude of a air-air reflected signal, as reference. The value of z_2 is calculated with the assumption of an approximately constant concrete density of $\rho=2.1.10^3$ kg/m^3.

The shear and compressional phase velocities c_s and c_p are evaluated using relations (6) and (4). In the following, the medium is considered as homogeneous. In this case the Young's modulus E and Poisson's ratio υ are given by

$$E = 2\rho c_s^2(1+\upsilon) \quad \text{and} \quad \upsilon = \frac{\left(1-2\frac{c_s^2}{c_p^2}\right)}{2\left(1-\frac{c_s^2}{c_p^2}\right)}. \tag{7}$$

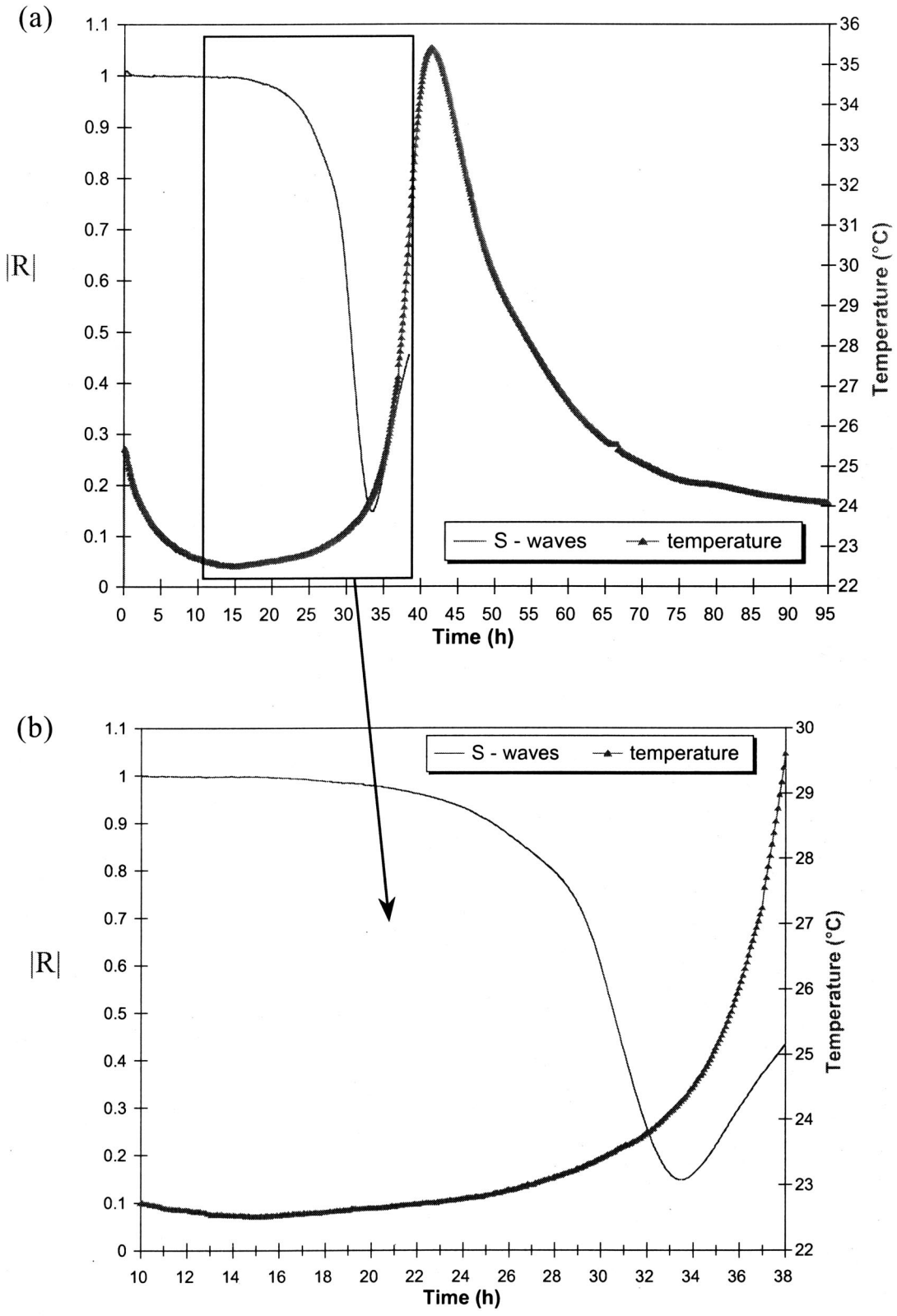


FIG. 2. Shear wave reflection coefficient and temperature in the sample between (a) 0 and 95 h (b) 10 and 38 h. The shear and compressional wave reflection coefficients lead to the calculation of the velocities and the elastic moduli.

It is clear that in its initial stage, the concrete has almost no shear elasticity, being fluid. The shear wave velocity is therefore zero. In consequence, the early values of the Young's modulus are zero, and the Poisson's ratio equals 0.5. The hardening will manifest itself by the increasing values of the Young's modulus and a decrease of the Poisson's ratio toward 0.2 approximately when the concrete has hardened.

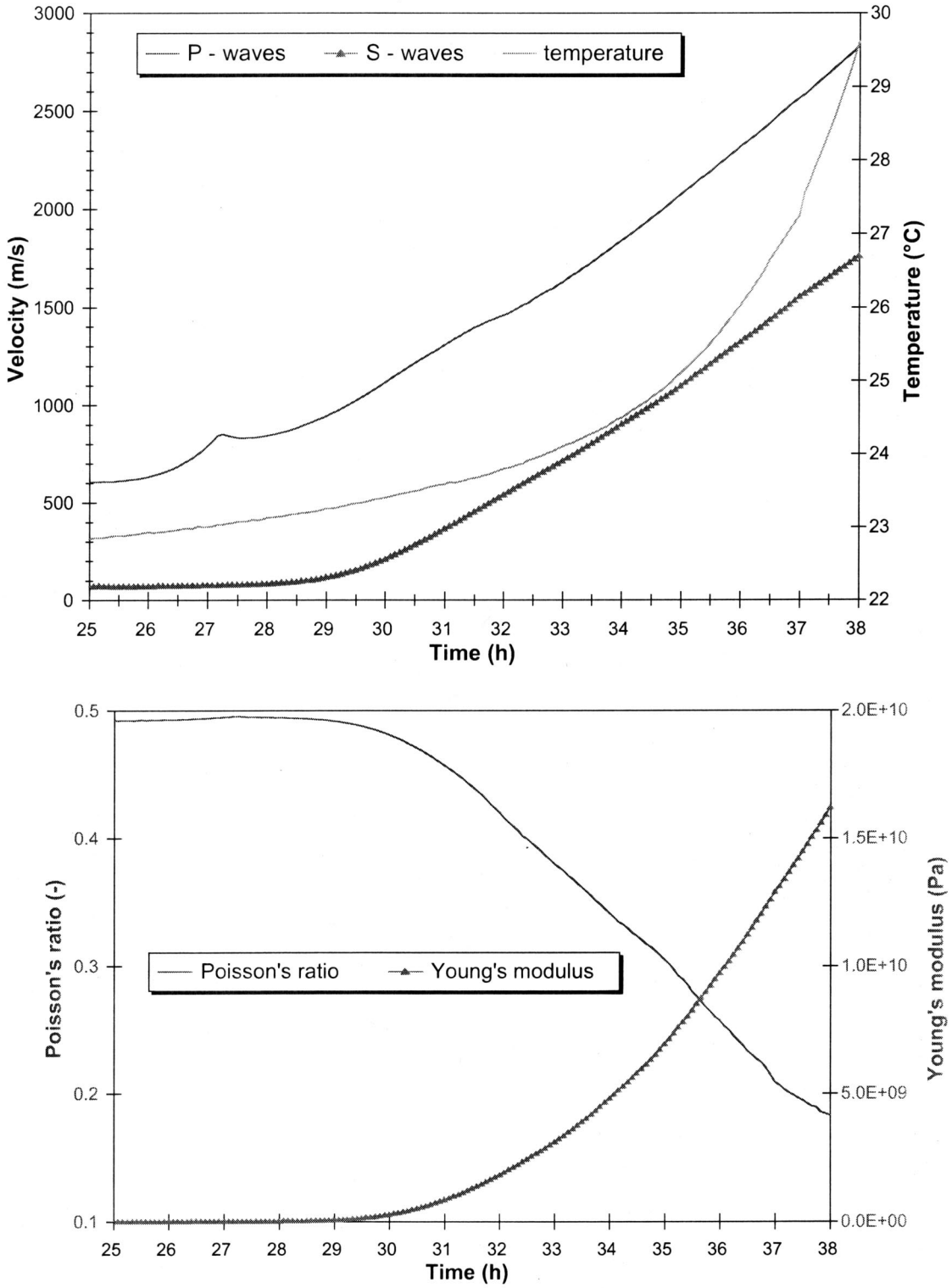


FIG. 3. (a) Velocities and temperature between 25 and 38 h. (b) Young's modulus and Poisson's ratio between 25 and 38 h.

B. Nonlinear elastic moduli

The nonlinear moduli are extracted by using the Preisach-Mayergoyz[6,7] space model of Guyer and McCall.[8–10] This model tells us that, in one dimension, the nonlinear contribution to the modulus can be written

$$k(x,t)=\beta\varepsilon+\ldots-\alpha[\Delta\varepsilon+\mathrm{sign}(\varepsilon)\varepsilon], \tag{8}$$

where β is the ''classical'' second order nonlinear parameter (Landau type[11]), and α is the ''nonclassical'' nonlinear parameter that describes hysteresis in the stress–strain relation. β and α can be calculated with the assumption that the measurements reflect a value proportional to acceleration,[12]

$$\beta\propto\frac{c_p^2A_2}{(A_1)^2}\quad\text{and}\quad\alpha\propto\frac{c_p^2A_3}{(A_1)^2}, \tag{9}$$

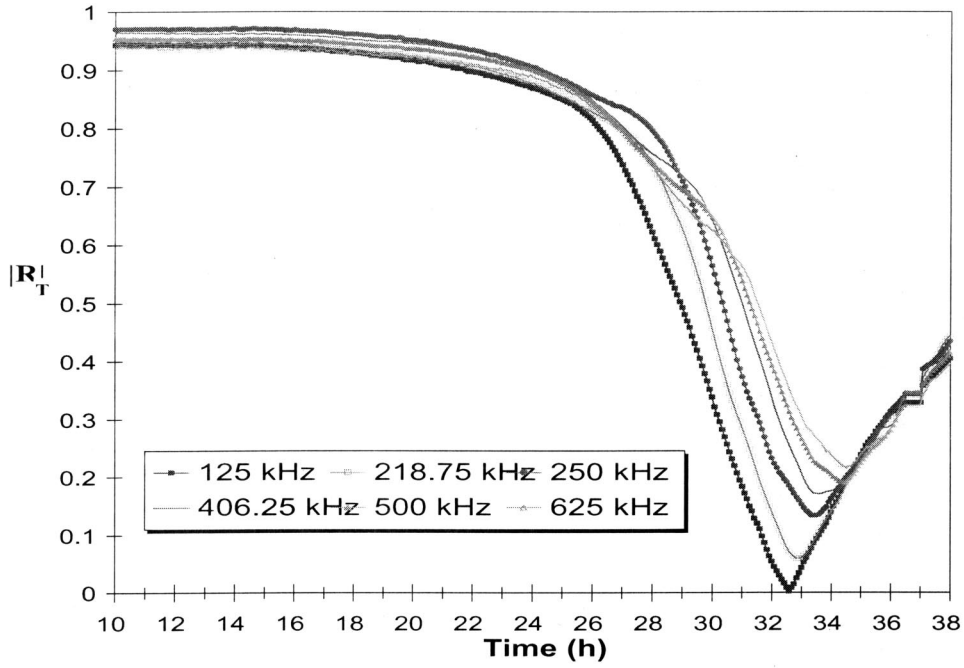


FIG. 4. Shear wave reflection coefficient for different frequencies.

where c_p is the compressional wave velocity, A_2 and A_3 are the amplitudes of the second and third harmonics, respectively, and A_1 is the amplitude of the fundamental (input wave), all measured at the detector. By plotting the fundamental wave amplitude versus the harmonic amplitudes at successively larger amplitudes, one can extract the nonlinear coefficients by fitting the data.

IV. RESULTS AND ANALYSIS

Figure 2 shows the temperature measured within the interior of the sample and the shear wave reflection coefficient measured at the base of the sample as described above, as a function of time during the experiment. Due to the progressive shrinkage of the material during curing, a disbonding of the transducers occurs at approximately 38 h after mixing (this is a problem we are currently addressing in order to study the behavior for longer periods of time). The primary chemical reaction begins at approximately 20 h as evidenced by the slope change in the temperature curve. Figure 3(a) shows the compressional (P) and shear (S) wave velocities, and the temperature, as a function of time between 25 and 38 h. Figure 3(b) shows Young's modulus and Poisson's ratio versus time, calculated from the reflection coefficient shown in the previous figure (Fig. 2). It is at this time that isolated solid regions begin to form in the material, however, the material remains in a primarily fluid state until about 29 h, as seen more clearly in the behavior of the moduli in Fig. 3(b) where the Young's modulus remains zero and the Poisson's ratio remains at 0.5. The correlations between the steep rise in temperature shown in Fig. 3 and the minimum in the reflection coefficient shown in Fig. 2 have been studied previously.[4] As seen in Fig. 4, a frequency effect is noticeable in the beginning of the cure starting at approximately 26 h. The acoustic impedance is larger at higher frequencies and the minimum shifts to later time with higher frequency. Full connection takes place at approximately 36–38 h slightly before the large thermal peak where the curves of the shear wave reflection coefficient for the different frequencies rejoin together (Fig. 4).

In the nonlinear portion of the experiment (Fig. 5), the transmission of the wave is possible only after 23 h, meaning that the fluid material entirely dissipates the wave before this time. The signal can be measured beginning at this time, but it is not until hour 28 that the second harmonic data are reliable, and at hour 30 for the third harmonic data (Fig. 5) corresponding to the rise of the Young's modulus from zero and the decrease of the Poisson's ratio from 0.5 (Fig. 3). We base our reliability on the power law dependence between the fundamental, and second and third harmonic amplitudes. According to Eq. (9) above from the *P-M* space model, the power law relation should be two in both cases. Noisy harmonic data with indefinable slopes are observed until the times noted above. Figure 5 illustrates several harmonic data sets taken at various times in the experiment. We show data sets taken during the fluid phase (bottom), during partial connection (top left) and when full connection takes place (top right). In all cases, at low drive levels the harmonics lie in the noise, and emerge at decreasing amplitude levels during the curing process. Figure 6(a) shows us only the reliable harmonic dependencies obtained throughout the experiment. In Fig. 6(a) are plotted the variation with time of the power law coefficient of the second and third harmonic amplitude dependences versus the amplitude of fundamental after 28 h. We observe a slope of approximately 2 beginning at hour 28 (A_{2f}) and 30 (A_{3f}). The data are noisy and the error bars indicate this, but the mean values are approximately 2. We believe that the observed dependence differs slightly from two because of the perturbation caused by resonances within the cell. As the experiments are being performed in a continuous sinusoidal excitation mode, a stationary wave within the cell is possible. This effect is less pronounced in the early part of the cure when the attenuation is more pronounced and

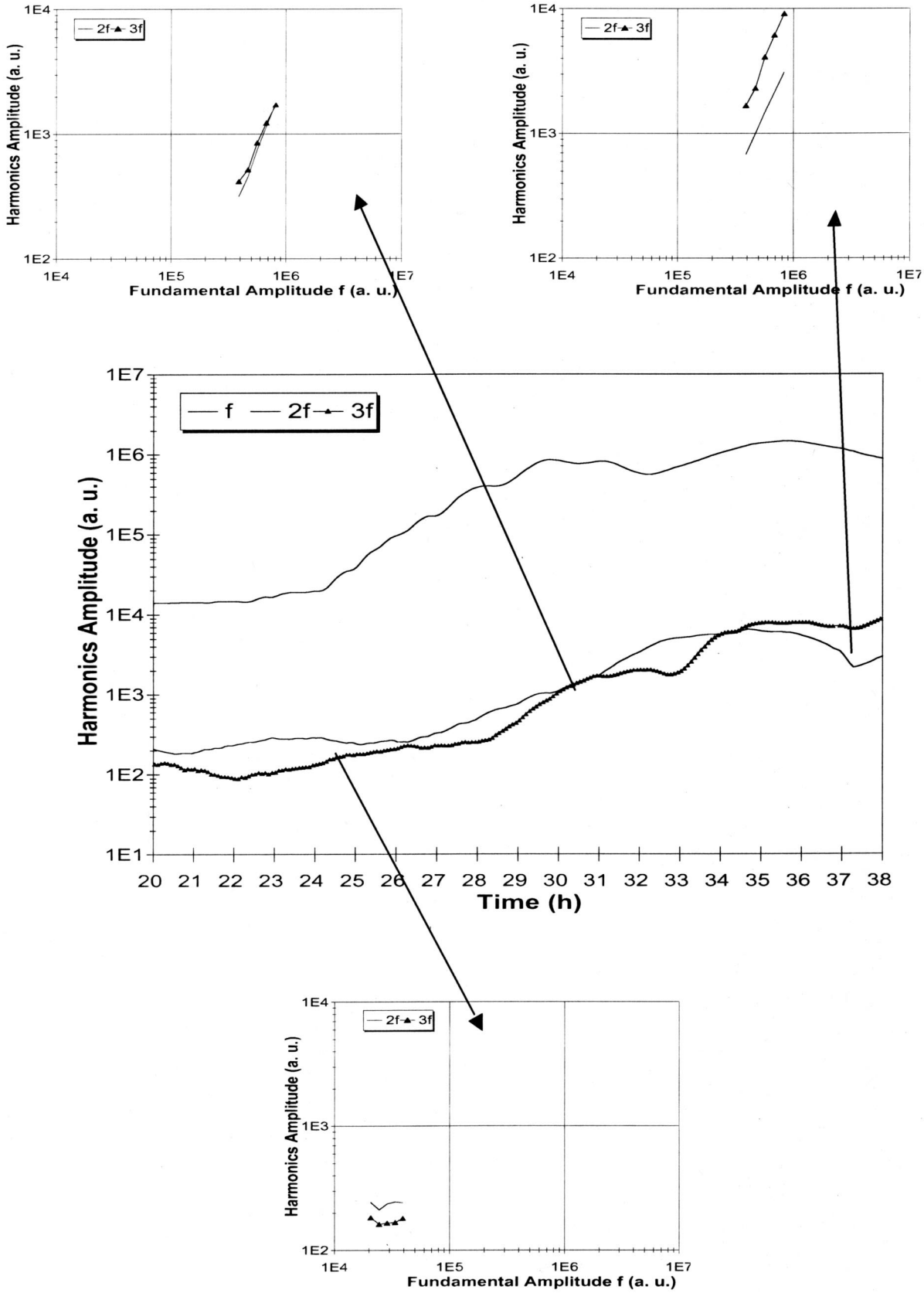


FIG. 5. Harmonic versus fundamental amplitudes.

the multiple reflected signals at the interfaces are small. As the attenuation decreases during the cure and the amplitude of the transmitted signal increases, the effect may become significant as can be seen on the oscillations on the amplitude of the fundamental. The slope of two in the harmonic amplitude dependence for the second and third harmonic indicates that we are in a nonlinear, nonclassical material. Classical nonlinear materials would show a third harmonic

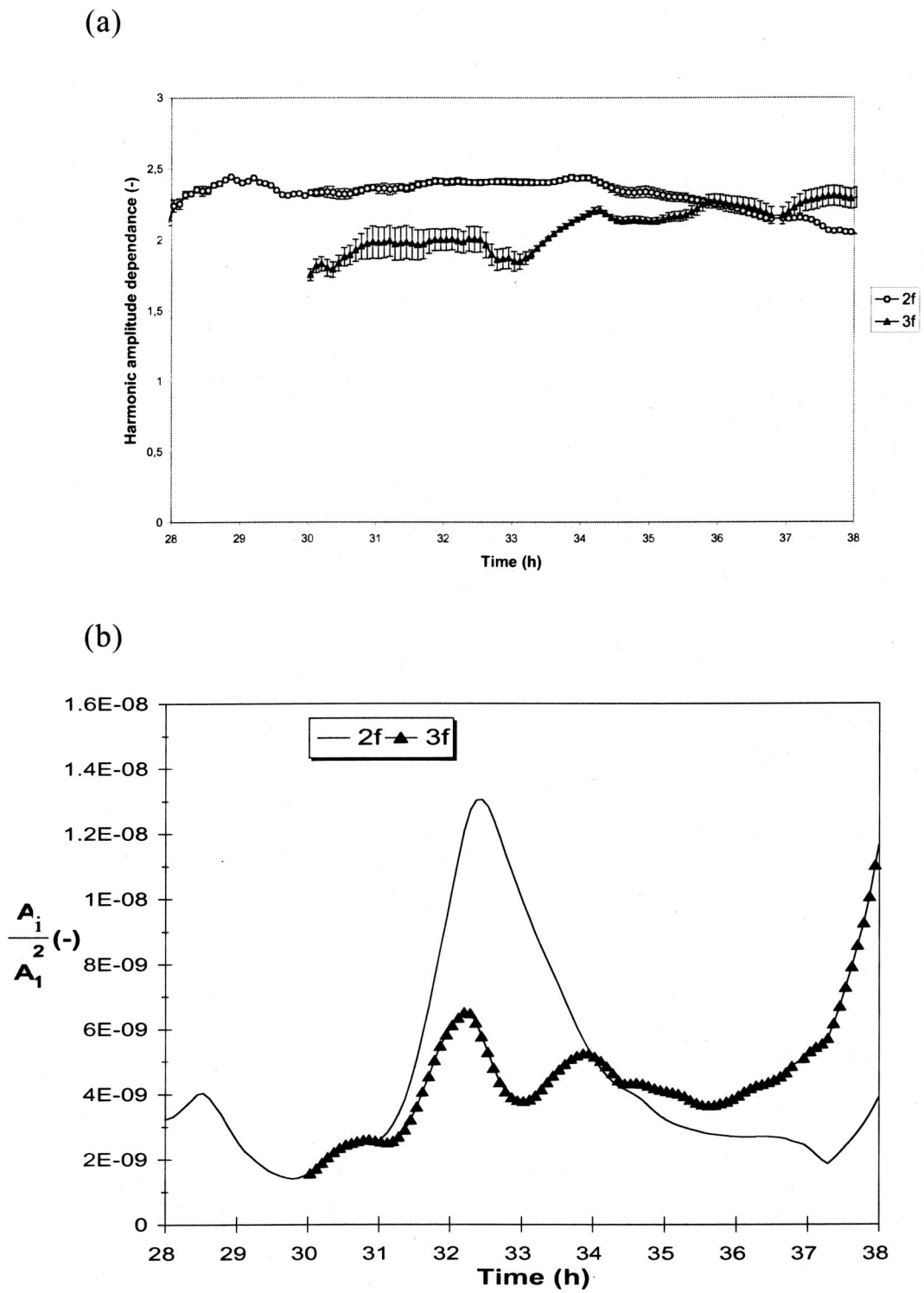


FIG. 6. (a) Harmonic amplitude dependence between 28 and 38 h. (b) Nonlinear coefficients between 28 and 38 h.

amplitude proportional to the cube of the fundamental amplitude. Thus we can estimate the nonlinear parameters $A_i/(A_1)^2$ of Eq. (9) that are proportional to nonlinear coefficients α and β of Eq. (8). The results are shown in Fig. 6(b). These are relative values of the nonlinear coefficients because we did not collect calibrated accelerations.

Thus, we can try to link the linear and nonlinear data during curing. From the monitoring of the reflection coefficient, we can see the changes in the mechanical state of the concrete with the different points in time in the thermal evolution like percolation point or points of connection of the different type of particles. These effects start at the percolation threshold that we assume to be at about 25 h. This point corresponds at the beginning of the thermal peak in the temperature curve [Fig. 2(a)] and to the beginning of a decrease in the shear modulus reflection coefficient in the shear wave reflection coefficient curves (Fig. 4) and in the harmonic amplitude versus time (Fig. 5) where we ob-

serve the beginning of the rise of the fundamental amplitude. During the thermal peak and the connection of particles in the material, we see a corresponding peak in the nonlinear parameters [Fig. 6(b)] that terminates at the point of the connection of the largest particles that is given by the linear part of the experiment corresponding to the coalescing of the reflection coefficient for different frequencies (Fig. 4).

V. CONCLUSIONS

By monitoring temperature, linear and nonlinear elastic properties, we can infer much regarding the evolution of the concrete during curing. The evolution of the reaction of hydration with the measure of the heat released informs us about the chemical behavior of the material in the experiment. The evolution of the reflection coefficient and moduli provide information about the mechanical state of the concrete including the liquid–solid phase change during the curing.

From monitoring harmonics we also see the phase change through the nonlinear coefficients. Before the thermal peak, in the liquid state, harmonics cannot be seen. These effects start at the percolation point where we notice the beginning of the harmonics generation and a nonclassical, nonlinear dependency. We observe a correlation between nonlinear response and linear response that relates to a known microstructure evolution of the connection of particles in the material. We relate the phase change with the thermal peak. We see that peaks in the nonlinear parameters linked to the nonlinear coefficients end at the point of the connection of the largest particles, as is given by the linear elastic results of the experiment (the reflection coefficient). The results indicate that the concrete is a nonclassical hysteretic elastic nonlinear material, like rock, some metals and ceramics.[13]

ACKNOWLEDGMENTS

We thank our colleagues, Koen Van Den Abeele, Robert Guyer, Alexander Sutin, and James TenCate for discussion and comments. This work was funded by CNRS (France), by the Franco-American Fulbright Commission, by institutional support (LDRD) at Los Alamos, and by the Office of Basic Energy Science of the US DOE. Portions of this work were presented in J. C. Lacouture, P. Johnson and F. Cohen Tenoudji, ''Linear and nonlinear investigating of concrete,'' International Congress on Acoustics, Rome, Italy, August, 2001.

[1] C. Vernet, J. Lukasik, and E. Prat, ''Nanostructure, porosity, permeability and diffusivity of ultra-high performance concrete,'' in *Proceedings of Intern. Symp. on HPC and RPC*, edited by P. C. Aïtcin *et al.* (RILEM Publisher, Sherbrooke, 1998), Vol. 3, pp. 17–35.
[2] G. De Schutter and L. Taerwe, ''Specific heat and thermal diffusivity of hardening concrete,'' Cem. Concr. Res. **25**, 593–604 (1995).
[3] A. Boumiz, C. Vernet, and F. Cohen Tenoudji, ''Mechanical properties of cement pastes and mortars at early ages: Evolution with time and degree of hydration,'' Adv. Cem. Based Mater. **3**, 94–106 (1996).
[4] V. Morin, F. Cohen Tenoudji, P. Richard, A. Feylessoufi, and C. Vernet, ''Ultrasonic spectroscopy investigation of the structural and mechanical evolutions of reactive powder concretes,'' *Proceedings of Intern. Symp. on HPC and RPC*, edited by P. C. Aïtcin *et al.* (RILEM Publisher, Sherbrooke, 1998), Vol. 3, pp. 119–126.
[5] P. A. Johnson, ''The new wave in acoustic testing,'' Mater. World **7**, 544–546 (1999).
[6] F. Preisach, Z. Phys. **94**, 277–302 (1935).
[7] I. D. Mayergoyz, ''Hysteresis models from the mathematical and control theory point of view,'' J. Appl. Phys. **57**, 3803–3805 (1985).
[8] K. R. Mccall and R. A. Guyer, ''Equation of state and wave propagation in hysteretic nonlinear elastic materials,'' J. Geophys. Res. **99**, 23887–23897 (1994).
[9] R. A. Guyer and K. R. Mccall, ''Hysteresis, discrete memory and nonlinear wave propagation in rock,'' Phys. Rev. Lett. **74**, 3491–3494 (1995).
[10] K. R. Mccall and R. A. Guyer, ''A new theoretical paradigm to describe hysteresis, discrete memory and nonlinear elastic wave propagation in rock,'' Nonlinear Processes in Geophysics **3**, 89–101 (1996).
[11] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, 3rd ed. (Pergamon, Oxford, England, 1986).
[12] K. E-A. VanDenAbeele, P. A. Johnson, R. A. Guyer, and K. R. McCall, ''On the quasi-static treatment of hysteretic nonlinear response in elastic wave propagation,'' J. Acoust. Soc. Am. **101**, 1885–1898 (1997).
[13] R. A. Guyer and P. A. Johnson, ''Nonlinear mesoscopic elasticity: Evidence for a new class of materials,'' Phys. Today **52**, 30–35 (1999).

An experimental study on antipersonnel landmine detection using acoustic-to-seismic coupling[a)]

Ning Xiang[b)] and James M. Sabatier
National Center for Physical Acoustics, University of Mississippi, 1 Coliseum Drive, University, Mississippi 38677

(Received 18 February 2001; revised 1 August 2002; accepted 2 September 2002)

An acoustic-to-seismic system to detect buried antipersonnel mines exploits airborne acoustic waves penetrating the surface of the ground. Acoustic waves radiating from a sound source above the ground excite Biot type I and II compressional waves in the porous soil. The type I wave and type II waves refract toward the normal and cause air and soil particle motion. If a landmine is buried below the surface of the insonified area, these waves are scattered or reflected by the target, resulting in distinct changes to the acoustically coupled ground motion. A scanning laser Doppler vibrometer measures the motion of the ground surface. In the past, this technique has been employed with remarkable success in locating antitank mines during blind field tests [Sabatier and Xiang, IEEE Trans. Geosci. Remote Sens. **39**, 1146–1154 (2001)]. The humanitarian demining mission requires an ability to locate antipersonnel mines, requiring a surmounting of additional challenges due to a plethora of shapes and smaller sizes. This paper describes an experimental study on the methods used to locate antipersonnel landmines in recent field measurements. [DOI: 10.1121/1.1543554]

PACS numbers: 43.28.En, 43.20.El [LCS]

I. INTRODUCTION

In recent years, the proliferation of landmines in many regions of the globe has focused worldwide attention on humanitarian landmine detection. Landmines keep refugees away from their homes and displaced away from their lands long after the guns of war fall silent. They endanger those who endanger no one, fostering fear in the innocent and young. This threat of antipersonnel (AP) landmines necessitates the development of humanitarian landmine detection systems. Current conventional detectors and ground-penetrating radars rely preferentially on detecting the metallic materials contained in landmines. The constant struggle to negate the advances of the other side has led manufacturers to mass produce nonmetallic mines. This requires novel approaches to mine detection. One of the more successful approaches involves the use of acoustic waves or mechanical vibrations for penetrating or exciting the ground surface.[1–8] While some of these approaches are still being investigated under laboratory conditions, this paper emphasizes field measurement results achieved using a laser Doppler vibrometer-based acoustic-to-seismic technique[1] to locate antipersonnel mines.

The first half-meter below an outdoor ground surface contains air-filled porous soil, which allows transmission of acoustic energy into the ground. Consequently, an acoustic signal impinging on the ground surface produces seismic motion of the surface.[9] This phenomenon has interested researchers since the 1950s[9–15] and is known as *acoustic-to-seismic coupling (A/S)*. The underlying physics of this seismic motion was identified as the result of motion of air in the pores of the soil. The motion of the air couples to the skeletal frame of the soil through momentum transfer and viscous drag at the porous walls, causing its energy to then be transferred to the soil frame.[9,13] This transfer of acoustic energy to the air-filled soil must be treated by a poroelastic wave model due to Biot,[10] for example. The Biot theory admits two compressional wave solutions. The waves are referred to as the waves of the first and second types (type I and type II[9,10]). The ground itself can be modeled as a two-dimensional poroelastic medium of depth d overlying a semi-infinite nonporous substrate. The air–soil interface can be assumed to be a free surface and the lower interface at depth d can be assumed to be in welded contact with an impermeable membrane between the two media. The air is allowed to flow across the upper boundary while it penetrates insignificantly below the elastic medium. For this physical system, wave equations with boundary conditions can be analytically solved (see Ref. 9 for more details). It is of practical significance in this context that this layered model explains a velocity response at the upper boundary due to the incident acoustic wave that would be the result of interference between the up- and down-going waves in the layer. In the following, we refer to the velocity response on the surface of the porous ground itself to the acoustic excitation as *background velocity*.

Typical burial depths of AP mines are usually only a few centimeters, placing them in this porous region of the soil. When an AP mine is buried, the A/S coupled waves undergoes distinct changes and can be sensed on the ground surface. Pertaining to the humanitarian landmine detection, it is also worthwhile mentioning that Donskoy[5] has shown that acoustic compliance of a mine is much greater than soils. He treated the mine–soil system as coupled harmonic oscilla-

[a)]Portions of this work have been presented in Proc. SPIE' 15th Conference on Detection and Remediation Technologies for Mines and Minelike Targets IV, edited by A. C. Dubey *et al.*, Orlando, FL, April, 2001.

[b)]Electronic mail: nxiang@olemiss.edu

tors. Buried landmines will resonate under an excitation by the A/S coupled energy. ''These resonances are due to the bending resonances of the mine casing's upper diaphragm.''[5]

Since the early 1980s, both theoretical and experimental studies of A/S coupling have been conducted.[11–15] Different types of sensors, including geophones and accelerometers, were used in the early measurements.[13–15] A review of early experimental investigations on A/S coupling for landmine detection can be found in Ref. 4. Geophones or accelerometers are contact sensors and therefore are less useful in landmine detection practice, because a safe detector requires noncontact, remote sensing. For this purpose, a feasibility study using a laser Doppler vibrometer (LDV) was conducted in the early 1990s.[16] The success of this study led to the development of an LDV-based acoustic mine detection technique.[1]

The acoustic technique has been successfully applied to outdoor detection of antitank (AT) mines[2] found in surrogate U.S. Army mine lanes. In a blind test for detection of AT mines in which the testers did not know the location of mines or even whether mines were present, the technique achieved a 95% probability of detection and $0.03/m^2$ false-alarm rate.[17] However, AP landmine detection is more challenging due to their smaller size and the variability in mine shape, size, and construction. This has led to alterations in equipment and techniques tailored to accomplishing the humanitarian mine-detection mission by locating AP mines. In this paper, Sec. II contains a description of the experimental configuration of an acoustic system for AP mine detection, Sec. III discusses recent field test results, and concluding remarks are presented in Sec. IV.

II. EXPERIMENTAL CONFIGURATION

A. Laser Doppler vibrometer

Safe mine detection requires noncontact remote sensing. Therefore, a laser Doppler vibrometer (LDV) has been used for sensing the A/S-coupled surface motion. This experimental study employed a single-point interferometer. The LDV emitted a laser beam onto the vibrating surface of the ground area under test. The surface vibration caused a Doppler frequency shift of the reflected laser light. A photodetector sensed the backscattered light from the measuring object coming along the opposite path back into the LDV (see Ref. 18 for more details about an LDV). The photodetector then emitted a frequency-modulated (FM) signal, which transmitted the surface velocity information. After FM demodulation of the detector, the output signal voltage was proportional to the instantaneous surface velocity of the vibrating point on which the laser beam was shining.

B. Measurement setup

For detection of AP mines, A/S coupling measurements were performed using a scanning LDV (PSV 200 manufactured by Polytec PI, Inc.). Figure 1 schematically illustrates the measurement setup. The LDV system was mounted between two subwoofer loudspeakers (Peavey 118 sub 8 HC) and over a third sound source (Altec model 290-4G) on a vibration-isolated platform mounted on a JCB 526 Loadall telescopic material handler. Because the LDV system, which

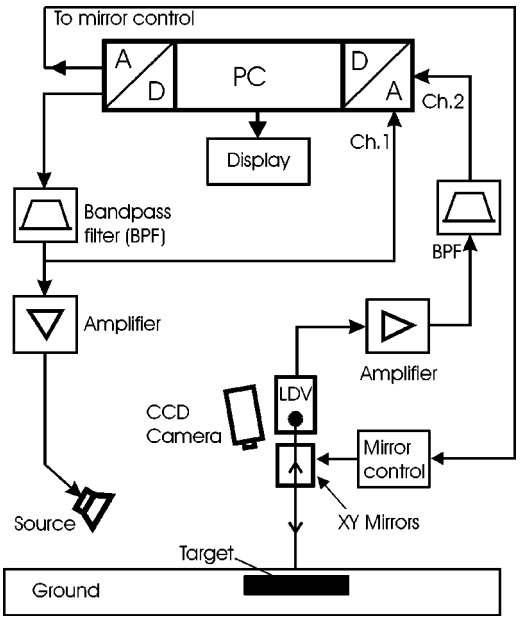


FIG. 1. Schematic diagram of the measurement system used in humanitarian landmine detection. There are two mirrors (X, Y) in the laser Doppler vibrometer to deploy the laser beam along the horizontal and vertical direction, respectively.

was equipped with a video camera and X–Y scanning mirrors (see Fig. 1), was in the sound field of the speakers, the LDV system was placed inside an isolation box. A PC monitor displayed a video image of the ground surface being scanned. Prior to scanning, a measurement grid was defined and superimposed on the image of the ground surface as shown in Fig. 2. On these images, the intersection points of the grid lines represent the exact scanning positions of the laser beam on the ground surface.

Experimental results using pseudorandom noise in the frequency range between 60 Hz and 10 kHz have revealed that the optimal frequency range for AP mine detection is between 100 and 680 Hz for the three outdoor surfaces considered. On the scanned patch of ground, the C-weighted sound-pressure level ranged between 90 and 110 dB. The LDV unit was placed inside the isolation box 2.3 m above the ground and the laser beam was focused onto the surface at an angle of 10 deg from normal to the road surface. The horn loudspeaker was suspended below the LDV platform as a sound source for the frequency range between 300 and 680 Hz. The center of the horn opening was placed approximately 1.8 m above the ground and 0.8 m from the center of a scanned patch. For the frequency range between 100 and 300 Hz, two subwoofers beside the LDV were used.

The sound source radiated periodic pseudorandom noise while the laser beam was deployed to predefined grid points one by one. In responding to the acoustic excitation, the instantaneous seismic velocity of the ground surface was sampled through one data collection channel, Fourier transformed, and averaged over several periods in a complex frequency domain. A resulting complex velocity function $\tilde{V}(f)$ was obtained at each grid point.

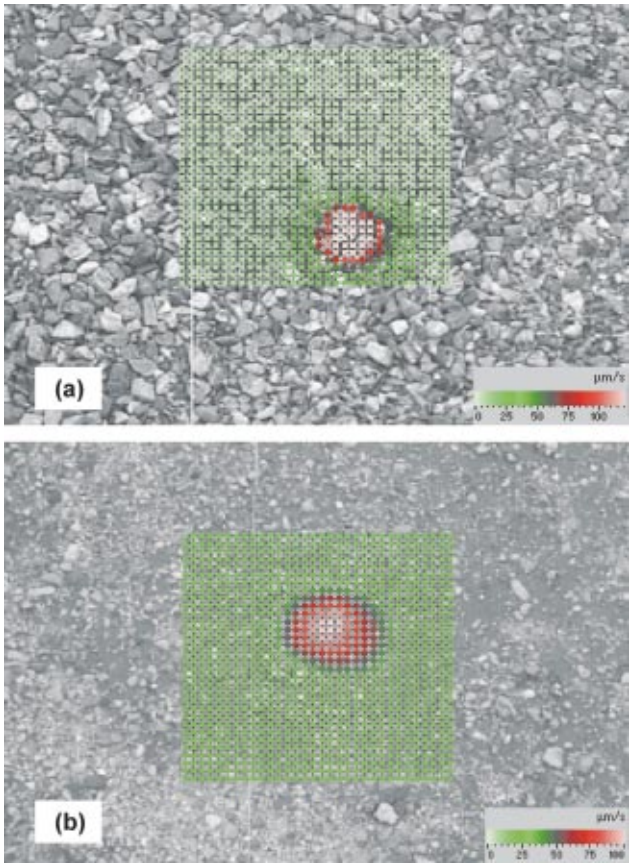


FIG. 2. Video camera photos of two road surfaces in test lanes. The video camera is integrated in the laser Doppler vibrometer system. The raster-scanning results in forms of color grid dots were superimposed onto the video image. In each image, the photo shows an area of 69 by 51 cm. A grid of 32 by 32 points covered an area of 30 by 30 cm. (a) Gray-gravel road surface with scanned color grid dots analyzed within a frequency range between 130 and 160 Hz. A PMA 3 antipersonnel mine was buried 2.5 cm deep in test lane 4. (b) Brown-gravel road surface with scanned color grid dots analyzed within a frequency range between 340 and 370 Hz. A VS 5.0 antipersonnel mine was buried 5.0 cm deep in test lane 5.

C. Field conditions

Real AP mines and mine simulants have been buried in the test lanes for more than 3 years. Real AP mines contained the original explosive material inside the mine case; however, the detonator was removed for safety purposes. Mine simulants were filled with nonexplosive substances with dielectric properties similar to explosive material. Data collections were performed during daylight hours. Three ground surfaces in which AP mines were buried have been investigated. Two of these surface [see the photographs in Figs. 2(a) and (b)] consisted of roads that were prepared by removing the existing soil and replacing it with one of two soil types. The third surface investigated was a natural soil [as shown in Fig. 6(a)]. Seismic refraction surveys, soil gradation, and sieve analysis[19] were carried out on these three soil materials. Table I lists relevant soil properties. Table II lists compression and shear wave speed. These data are used to characterize the lanes and useful for future modeling.

D. Basic analysis method

The magnitude of the spectrum of the velocity function $\tilde{V}(f)$ at each grid point was integrated over a frequency band chosen according to the occurrence of consistent amplifications in magnitude velocity in the presence of a mine

$$M_{ij}\int_{f_1}^{f_2}|\tilde{V}_{ij}(f)|df, \tag{1}$$

with f_1, f_2 denoting the lower and upper frequency limits, respectively. i,j was the subscript of a grid point on the ith row and jth column of the grid. In this way, a single-valued magnitude velocity could be presented as data points on a color dot map.

An example of a scanning result is shown in form of a color dot map in Fig. 3(a), which is a cut from Fig. 2(a). This scanning result was obtained from field measurements on a gray-gravel road. Colors were automatically assigned proportionate to the integrated values of magnitude velocity in the map. The map was superimposed onto the video image of the ground surface. The circle on the color dot map indicated the mine location. Based on integrated magnitude values at individual scanning points, processing in terms of nearest-neighbor and spatial filtering using two-dimensional median filter[20] yielded a smoothed color map as shown in Fig. 3(b). Its 3D presentation is illustrated in Fig. 3(c).

TABLE I. Soil properties of three test lanes.

Lane	% Gravel	% Sand	% Fines	Field density	Porosity	Flow resistivity range (g.s.cm^{-3})
Lane 4	56.0	36.0	8.0	1.54 g.cm-3	41.89%	30.64–274.88
Lane 5	21.4	56.7	21.9	1.48 g.cm-3	44.15%	283.53–1430.12
Natural soil	0.5	67.1	32.4	1.64 g.cm-3	38.11%	Not available

III. DISCUSSION OF MEASUREMENT RESULTS

Using the measurement setup and the analysis method described in the previous section, field measurements and subsequent analysis were performed on buried AP mines at the test lanes (see Table I). This section discusses some relevant issues of the LDV-based A/S coupling AP mine detection based upon the results achieved from the field tests.

A. Detection of buried landmines

Figure 4 shows several magnitude spectra of velocity functions on and off the target for the PMA 3 AP mine that is 10 cm in diameter, buried 2.5 cm deep at test lane 4 (listed in

TABLE II. Compression and shear wave speed of three test lanes.

	Compression wave (P-wave)		Shear wave (S-wave)	
Lane	Layer speed (m/s)	Averaged depth (m)	Layer speed (m/s)	Averaged depth (m)
Lane 4	274	0	84	0
	526	0.3	150	0.11
	716	0.82	260	0.33
Lane 5	320	0	90	0
	533	0.1	260	0.27
	693	0.78	340	1.7
Natural soil	259	0	38	0
	472	0.06	274	0.15
	693	0.6	⋯	⋯

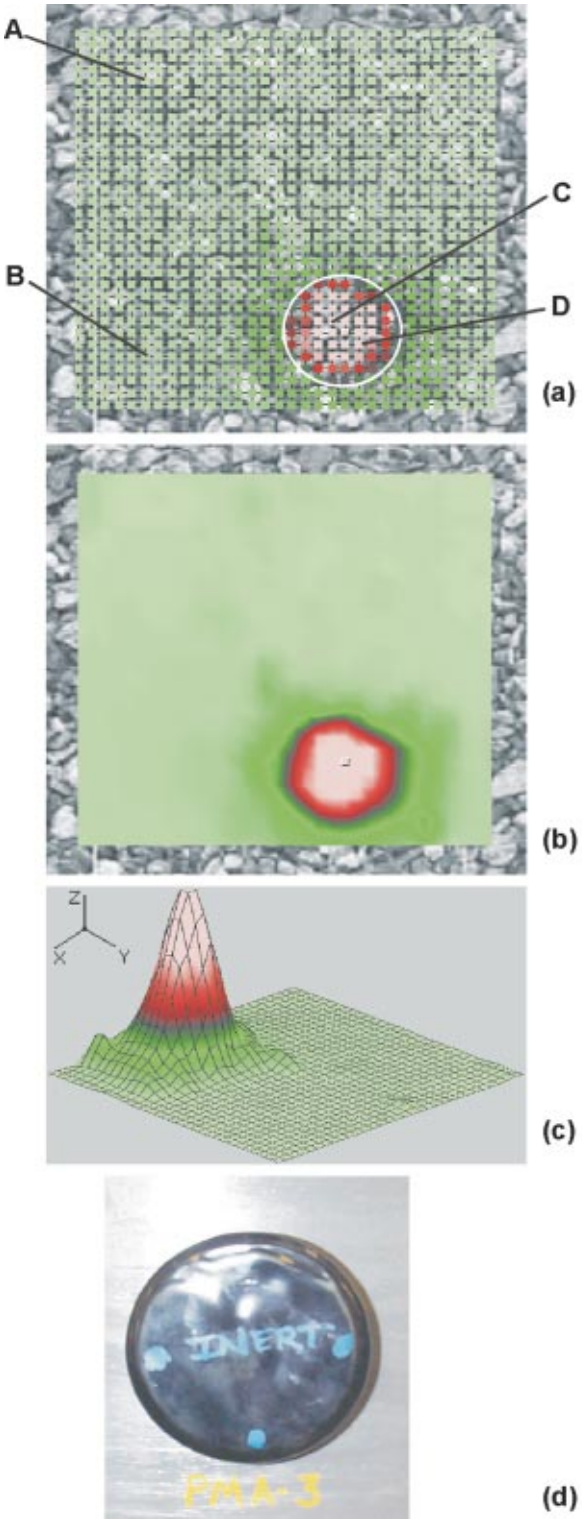


FIG. 3. A grid of 32 by 32 points was defined by covering a scanned patch of 30 by 30 cm (spatial resolution: 0.97 cm) in lane 4. A PMA 3 antipersonnel mine with a circular shape was buried 2.5 cm deep. The integrated velocity value in the frequency range between 130 and 150 Hz is presented. (a) Color dots. The circle indicates the target location. On points A, B, C, and D individual measured magnitude spectra were illustrated in Fig. 4. (b) Color map achieved from (a) in terms of image processing. (c) Three-dimensional color presentation derived from (b). (d) Photograph of a plastic PMA 3 antipersonnel mine that was 10 cm in diameter.

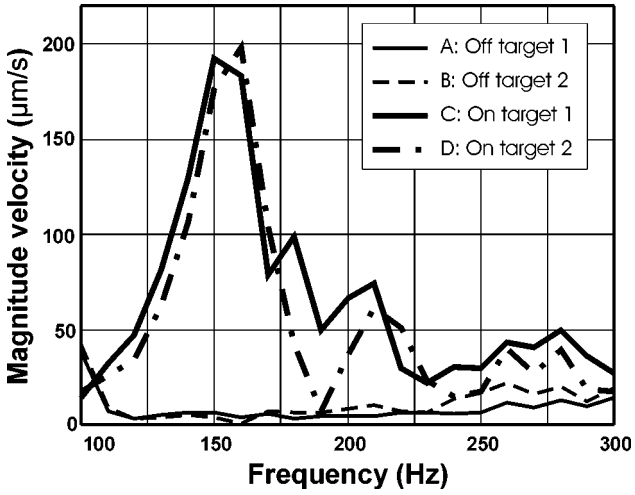


FIG. 4. Magnitude velocity spectra evaluated from the scanning result shown in Fig. 3. Individual magnitude velocity spectrum evaluated from the grid points A, B, C, and D indicated in Fig. 3(a).

Table I). Frequency resolutions of 2.5–10 Hz were often used to represent the discrete velocity functions $\tilde{V}(f_i)$. The magnitude spectra on top of the target indicate that the A/S coupled energy excites resonant vibrations in the buried AT landmines (see also Refs. 5 and 6). In Figs. 3(a) and (b), a region of interest could be identified in both presentations, indicating the presence of a mine. The size of the mine could be estimated by counting grid points [in Fig. 3(a)] across the region. When a region of interest was thought to indicate a mine, the middle of the region was assumed to be associated with the center position of the buried mine. The laser beam could then be moved to a point in that region, thereby marking the location on the ground. It usually pointed to the center of the mine with a radial accuracy of less than 2 centimeters. Accuracy was significantly enhanced through the use of closer grid-point spacing.

Mine types, burial depth, and ground surfaces predominated in determining A/S coupling responses of on-target velocity. Field measurements have shown that the maximum ratio between on- and off-target velocity in a suitable frequency band could range from multiples of 10 down to 2.

B. Background velocity

Road types (depending on the construction material), deep ground layering, and weathering conditions influence A/S coupling responses of off-target velocity, henceforth referred to as background velocity responses. As shown in Fig. 4, the detection of mines depends upon exploiting the difference between the ground velocity over a mine and away from a mine. Therefore, evaluations of background velocity functions are important for detecting landmines. Using the brown-gravel road at test lane 5 as an example, Fig. 5 illustrates background velocity functions both within a single 0.36 m^2 (0.6×0.6-cm) patch and across six similar patches distributed randomly along the road. The standard deviation

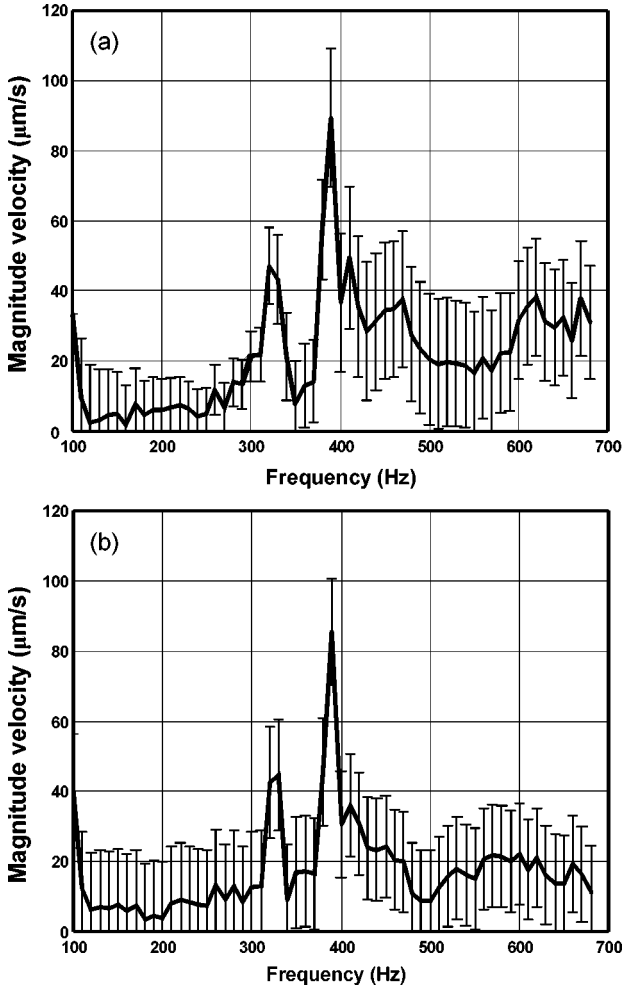


FIG. 5. Background magnitude velocity spectra in the brown-gravel road. The standard derivation of the velocity is shown as error bars for each frequency. (a) Averaged background magnitude velocity spectrum evaluated from five individual spectra. The individual background magnitude velocity spectra are randomly selected within a single 60- by 60-cm patch. (b) Averaged background magnitude velocity spectra evaluated over six (60- by 60-cm) patches.

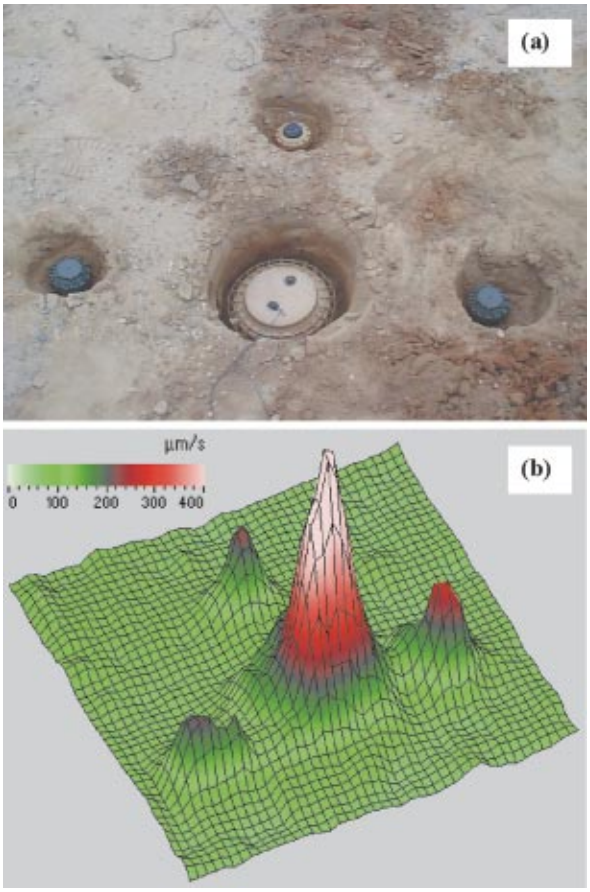


FIG. 6. Scanning result on a combination of antitank and antipersonnel landmines. A plastic VS 2.2 antitank mine, 24 cm in diameter, was buried 6 cm deep surrounded by three plastic antipersonnel mines (one TS 5.0 mine on opposite sides of the antitank mine and one VS 5.0 on a third side). These three antipersonnel mines, 9 cm in diameter, were buried 3 cm deep. A grid of 49 by 49 points covering an area 1.1 by 1.1 m was defined, resulting in a spatial resolution of 2.3 cm. (a) Relative positions of the mines before burial in the natural soil. (b) The scanning results in a three-dimensional presentation. Magnitude spectra were integrated within the frequency range between 100 and 300 Hz.

of the velocity is shown as error bars for each frequency. At a given acoustic excitation in a range around a C-weighted sound level of 100 dB, the averaged velocity assumes lower values of 6–10 μm/s in the frequency range between 100 and 300 Hz and higher values of about 15–30 μm/s between 450 and 680 Hz. One sharp peak at 320 Hz has been identified as being caused by the LDV mirror resonance frequency. Another sharp peak at 390 Hz has been identified as emanating from the first room mode of the sound isolation box covering the LDV. When a number of scanned points on the ground assume clearly higher velocity values than the background velocity, the differences between on- and off-target velocities are mapped on the scanning image as regions of interest.

C. Spatial resolution

The spatial resolution of scanning is of vital importance for mine detection, especially since the variability of mine sizes and shapes is relatively large among AP mines. Based on finding the consistent amplifications of the magnitude velocity over a certain frequency range, an adequate spatial resolution under ideal field conditions has to obey the spatial Nyquist sampling principle. This requires that at least two grid points along the minimum dimension on top of a mine be included. Moreover, a spatial oversampling protocol allows visual recognition of target shape, estimation of the target size, and, therefore, can distinguish mines from background clutter. Clutter may result in high magnitudes of velocity attributable to such causes as inhomogeneity of the ground. Figure 3 illustrates scanned results with a spatial resolution of 1 cm. As expected, the target images are circular in shape due to spatial oversampling. Clutter often causes irregular-shaped and smaller-sized images in the color maps. The spatial oversampling, however, results in longer detection time.

Figure 6 shows a combination of antitank (AT) and AP mines. A VS 2.2 AT mine, 24 cm in diameter, was buried 6 cm deep surrounded by three AP mines (two TS 5.0 mines on opposite sides of the AT mine and one VS 5.0 on a third side). These three AP mines, 9 cm in diameter, were buried 3 cm deep. All these four mines were made of plastic. Figure

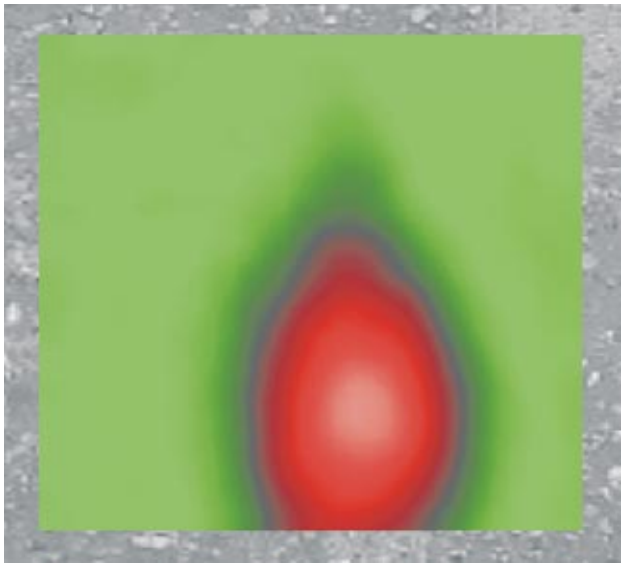


FIG. 7. Scanning results in form of a color map on a PMD 6 antipersonnel mine buried 5 cm deep. Its rectangular shape has a (top view) length of 20.5 cm and a width of 9 cm. A grid of 32 by 32 points covering an area 30 by 30 cm was defined, resulting in a spatial resolution of 1 cm. Magnitude spectra were integrated within 280–310 Hz.

6(a) shows the relative positions of the mines before burial. To resolve an image of the smaller size of AP mines, a grid of 49 by 49 points covering an area 1.1 by 1.1 m was defined, resulting in a spatial resolution of 2.3 cm. Figure 6(b) illustrates the scanning results in a three-dimensional presentation. Magnitude spectra were integrated within the frequency range between 100 and 300 Hz. As one can see, one AT mine surrounded by three AP mines could be clearly resolved with the chosen spatial resolution.

Figure 7 shows results scanned on a PMD 6 AP mine buried 5 cm deep. Its case is a wooden box. Its rectangular plan form has a (top view) length of 20.5 cm and a width of 9 cm. A grid of 32 by 32 points covering an area 30 by 30 cm was defined, which resulted in a spatial resolution of 1 cm. Magnitude spectra were integrated within a frequency range between 280 and 310 Hz. The mine image presented an elongated shape even though its rectangular shape could not be resolved by the chosen spatial resolution.

Figure 8(a) shows results for a VAL 69 AP mine buried 5 cm deep. Figure 8(b) shows a photograph of the mine. In the plan view, the diameter is 14 cm. A grid of 32 by 32 points covering an area 30 by 30 cm was defined, which resulted in a spatial resolution of 1 cm. Magnitude spectra were integrated within a frequency range between 360 and 400 Hz. Due to the irregular shape of the top of this mine, the scanned results tended to show an irregular image.

D. Frequency resolution and analysis

A/S coupling responses for AP mines often presented amplifications over a broad frequency range, as the example in Fig. 4 shows. The integration of the magnitude velocity spectrum for each scanning grid point was performed within a narrow frequency band using Eq. (1). A color map representing the integrated magnitude velocity values over the entire grid was then formed. This narrow-band procedure

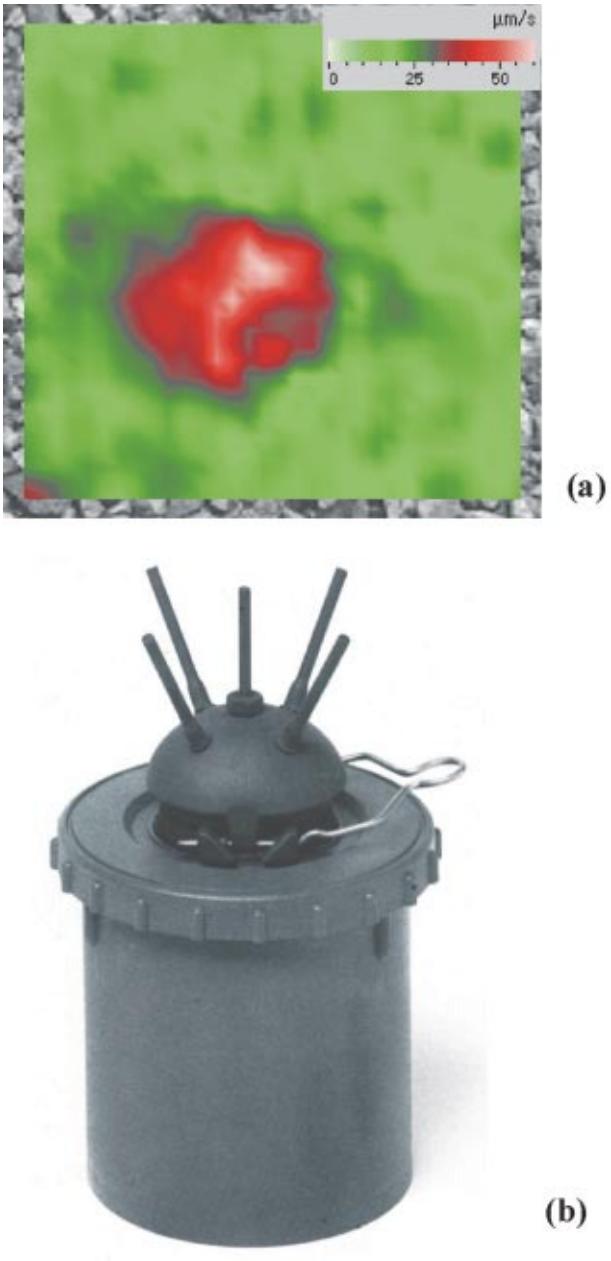


FIG. 8. (a) Scanning results in the form of a color map. A grid of 32 by 32 points was defined covering an area of 30 by 30 cm resulting in a spatial resolution of 0.97 cm. Magnitude spectra were integrated between 360 and 400 Hz. Due to irregular shape at the top of this mine, the scanned results tended to show an irregular image as well. (b) Photograph of a VAL 69 antipersonnel mine. The largest circular diameter is 14 cm.

was repeated by stepping through the entire frequency range with an overlap from one frequency band to the next. Figure 9 illustrates the narrow-band analysis of scanning results obtained on a VS 5.0 mine buried 5 cm deep in test lane 5. By stepping a narrow frequency band with a 30-Hz bandwidth through the frequency range between 240 and 370 Hz, the mine image remained in the same position with almost the same size, while a smaller clutter image could only be seen in the band between 280–310 Hz and 340–370 Hz.

The narrow-band analysis was based on an adequate frequency resolution (spacing) in the data. Higher frequency resolution or smaller frequency spacing facilitated broader possibilities to optimize the narrow-band analysis. The con-

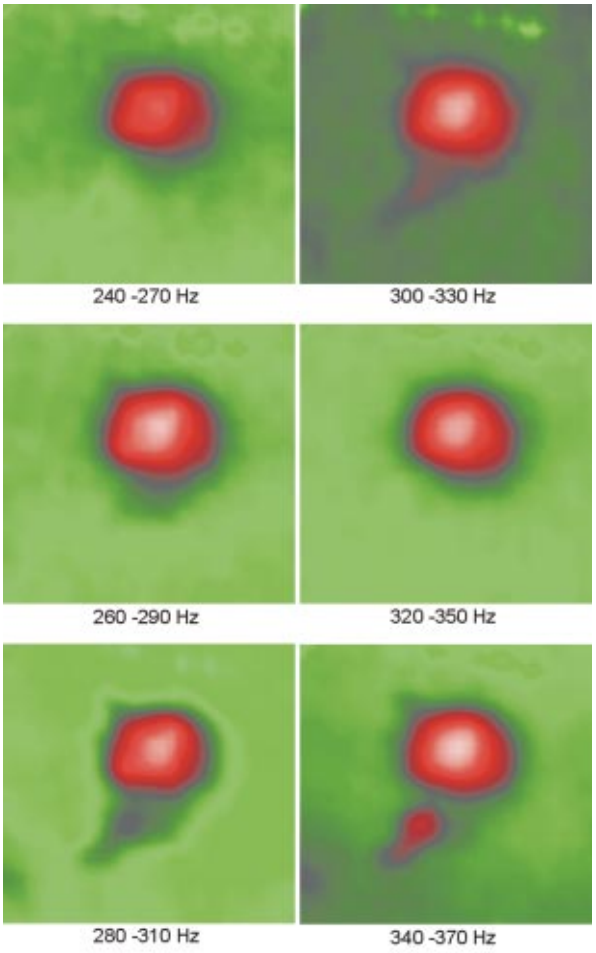


FIG. 9. Narrow-band analysis of scanning results obtained on a VS 5.0 mine buried 5 cm deep in the brown-gravel road. By stepping a narrow frequency band with a 30-Hz bandwidth through the frequency range between 240 and 370 Hz, the mine image remained in the same position with almost the same size, while smaller clutter could only be seen in the band between 280–310 Hz and 340–370 Hz.

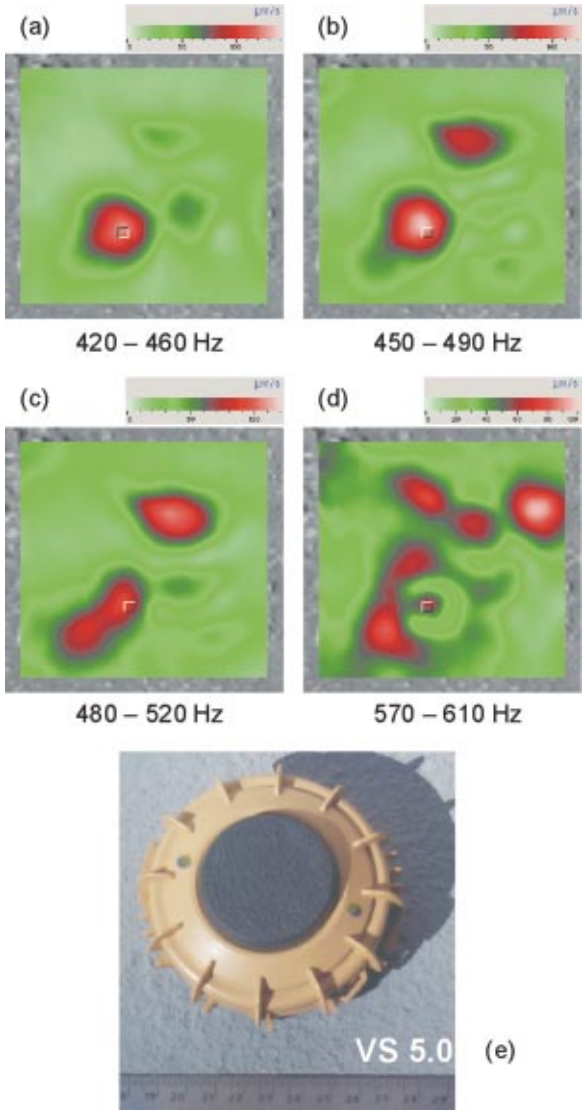


FIG. 10. Scanning results in form of color maps on a VS 5.0 mine buried 5 cm deep in the brown-gravel road, in the frequency range between 420 and 610 Hz with a frequency resolution of 5 Hz. A grid of 42 by 42 points covering an area of 50 by 50 cm resulted in a spatial resolution of 1.2 cm. Within this frequency range, there are two regions of interest. One region on the lower-left side was identified as mine response. The region on the upper-right side with an irregular shape was due to clutter.

sistency in the position and the size of a target, while stepping a narrow-band filter through a relative broad frequency range, enhanced distinguishing mines from background clutter. Higher frequency resolution, however, required collecting a larger number of data points. In the experimental study we chose a resolution in order of 2.5–10 Hz that had to be predefined for the data collection. Both the spatial and frequency resolution determined the speed of scanning. Taking an example as shown in Fig. 3, 32 by 32 scanning points covered an area of 30 by 30 cm. A frequency resolution of 10 Hz was used at a sampling frequency of 1.024 kHz, yielding a signal period of order of 100 ms. Using three averages the scanning took about 5–10 min. A detection system with multiple laser beams and parallel processing will increase the scanning speed.

E. Clutter responses

Loose soil and inhomogeneities at a shallow depth in the ground can lead to high magnitude in the A/S coupling responses. Figure 10 illustrates scanning results for an VS 5.0 mine buried 5 cm deep in test lane 5. In the frequency range between 420 and 460 Hz the mine appears as a single circular region with two other clutter images in smaller amplitude. In the frequency between 450 and 520 Hz, two regions of interests could be found. The region on the lower-left side was associated with the mine response. The more irregular-shaped region on the upper-right side was clutter. Figure 10 demonstrates difficulties in distinguishing mines from clutter since clutter can result in similar magnitude responses in the same frequency ranges with comparable sizes to a mine [see Figs. 10(c) and (d)]. When stepping the narrow-band filter to 570–610 Hz, a ring-form pattern caused by the buried AP mine appeared in the map [Fig. 10(d)]. The ring-form pattern happens at a frequency pointed out by Yu *et al.*[22] where the soil–mine system[5] shows an ''antiresonance'' at which the surface vibration of certain regions on top of the mine becomes lower than those around the mine. Figure 11 illustrates representative magnitude spectra on and off the target. The magnitude spectra on the top of the mine show distinct

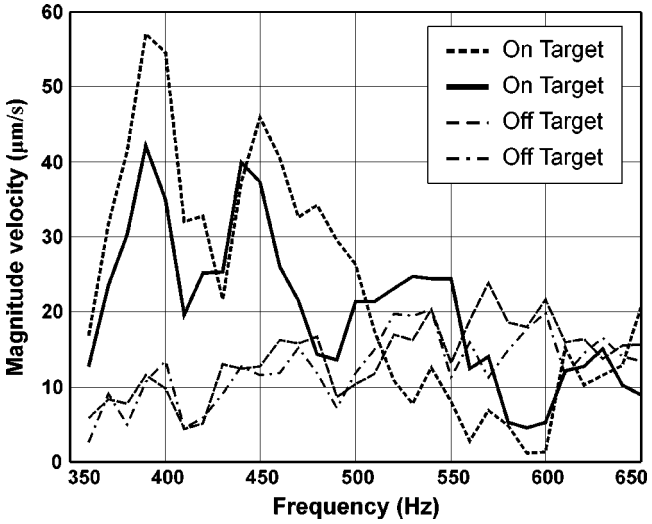


FIG. 11. Magnitude velocity spectra scanned in the brown-gravel road where a VS 5.0 AP mine was buried 5 cm deep. In the frequency range between 580 and 605 Hz the magnitude spectra on the top of the mine become clearly lower than those away from the mine. In this frequency range, the mine manifests itself as a ring-form pattern in the magnitude velocity map [as shown in Fig. 10(c)].

resonant behavior in the frequency range between 350 and 410 Hz as well as 430 and 480 Hz. In addition, the on-target magnitude spectra become so small to be clearly lower than those away from the target (background) in the antiresonance frequency range between 580 and 605 Hz. These phenomena cannot be observed on the regions of the clutter.

This section has discussed some relevant issues about the AP mine detection based on the field test results. In order to perform mine detection using A/S coupling, the surface velocity on an AP mine should be clearly higher than the background velocity within a broad frequency range. Field measurements have shown that the maximum ratio between on- and off-target velocity in a suitable frequency band ranged from several multiples of 10 down to 2 due to the mine resonance. A spatial oversampling supported the recognition of target shape and size and thus enhanced the detectability of AP mines. Narrow-band frequency analysis has been found to be a good tool to enhance the detectability. When stepping a narrow-band filter through the overall frequency range, mine images remain relatively stable while clutter often causes high variability in object shape, size, and position. The resonance and antiresonance behavior of the mine–soil system[22] can also be exploited to help distinguish mines from clutter.

IV. CONCLUDING REMARKS

The porous nature of the ground permits landmine detection based upon the distinct changes of the acoustic-to-seismic coupled motion on the ground surface. This led to development of a new technique for humanitarian landmine detection: the laser-Doppler vibrometer-based acoustic landmine detection system. The current acoustic landmine detection system utilized a scanning single-beam laser Doppler vibrometer for sensing the acoustic-to-seismic coupled motion on the ground surface. Complex surface velocity functions of frequency were measured using a remote, raster-scanning technique. In general, the magnitude of the velocity function was used to detect antipersonnel landmines. The scanning results were evaluated within a frequency range between 100 and 680 Hz. Color maps were formed to image mine location, size, and shape. Landmine responses were found to be broadband in nature. A landmine was determined to be present when there was consistent amplification of the magnitude velocity over a relatively broad frequency band and when a circular shape in the scanning image remained intact when stepping through the overall frequency range with a narrow-band filter. Often clutter entered and exited the image as the narrow bands were sequentially changed. Sometimes clutter was also broadband. In these cases, the size of the region of interest was exploited for a decision in addition to the circular shape. For humanitarian landmine detection, the acoustic system has demonstrated a capability for detecting antipersonnel landmines buried in the subsurface of the ground with a high probability of detection. The primary challenge is to maintain a low false-alarm rate while retaining this high probability of detection. In the frequency range in which most antipersonnel mines result in amplified magnitude velocity spectra, some clutter and anomalies can also appear. In addition to the shape and size criteria, the AP landmine detection relies on recognition and identification of the resonant behavior of the soil–mine vibration system to distinguish them from the clutter and anomalies. Extensive collection of field results on broad ranges of burial AP landmines along with soil types and systematic analysis of mine signatures will promote our knowledge on enhancing the detection ability.

Up to the current time, the research effort has focused on demonstrating the feasibility of this technique in field measurements rather than achieving a high scanning speed. In the near future, efforts are expected to model the physics of the soil–mine vibration system excited by the acoustic/seismic coupling,[23] to collect and identify unique acoustic-to-seismic coupling signatures of antipersonnel landmines, systematically analyzing clutter responses. Efforts will also be made to increase the speed[24] and detection performance[25] of the current technique and to automatically recognize targets.[26]

ACKNOWLEDGMENTS

This work is supported by the U.S. Army Communications-Electronics Command Night Vision and Electronics Sensors Directorate. The authors are grateful to Thomas Witten and Erik Rosen for their assistance in this effort. Their thanks also go to Daniel Costley for construction of the LDV isolation box and Ronald Craig for his dedication and expertise in collecting the data for this paper. The authors would like to express their thanks to the Associate Editor Mr. Louis C. Sutherland for his encouragement and constructive comments.

[1] J. M. Sabatier and K. E. Gilbert, ''Method for detecting buried object by measuring seismic vibrations induced by acoustical coupling with a remote source of sound,'' U.S. Patent, No. 6,081,481 (2000).
[2] J. M. Sabatier and N. Xiang, ''An investigation of acoustic-to-seismic coupling to detect buried anti-tank landmines,'' IEEE Trans. Geosci. Remote Sens. **39**, 1146–1154 (2001).

[3] J. Sabatier and N. Xiang, ''Laser-Doppler Based Acoustic-to-seismic detection of buried mines,'' Proc. SPIE **3710**, Conference on Detection and Remediation Technologies for Mines and Minelike Targets IV, edited by A. C. Dubey *et al.* (1999), pp. 215–222.
[4] N. Xiang and J. Sabatier, ''Land mine detection measurements using acoustic-to-seismic coupling,'' Proc. SPIE **3710**, Conference on Detection and Remediation Technologies for Mines and Minelike Targets V, edited by A. C. Dubey *et al.* (2000), pp. 645–655.
[5] D. Donskoy, A. Ekimov, N. Sedunov, and M. Tsionskiy, ''Nonlinear seismo-acoustic land mine detection and discrimination,'' J. Acoust. Soc. Am. **111**, 2705–2714 (2002).
[6] W. R. Scott, Jr., J. S. Martin, and G. Larson, ''Experimental model for a seismic landmine detection system,'' IEEE Trans. Geosci. Remote Sens. **39**, 1155–1164 (2001).
[7] S. W. McKnight, C. A. DiMarzio, W. Li, D. O. Hogenboom, and G. O. Sauermann, ''Laser-induced acoustic detection of buried objects,'' Proc. SPIE, Vol. 3392, Conference on Detection and Remediation Technologies for Mines and Minelike Targets III, edited by A. C. Dubey *et al.* (1998), pp. 231–238.
[8] C. G. Don, D. E. Lawrence, and A. J. Rogers, ''Using acoustic impulse to detect buried objects,'' Proc. ICA 1998, pp. 1759–1760.
[9] J. M. Sabatier, H. E. Bass, L. N. Bolen, K. Attenborough, and V. V. S. S. Sastry, ''The interaction of airborne sound with the porous ground. The theoretical formulation,'' J. Acoust. Soc. Am. **79**, 1345–1352 (1986).
[10] M. A. Biot, ''Theory of propagation of elastic waves in a fluid saturated porous solid. II. Higher frequency range,'' J. Acoust. Soc. Am. **28**, 179–191 (1956).
[11] H. E. Bass, L. N. Bolen, D. Cress, J. Lundien, and M. Flohr, ''Coupling of airborne sound into the earth: Frequency dependence,'' J. Acoust. Soc. Am. **67**, 1502–1506 (1980).
[12] K. Attenborough, J. M. Sabatier, H. E. Bass, and L. N. Bolen, ''The acoustic transfer function at the surface of a layered poroelastic soil,'' J. Acoust. Soc. Am. **79**, 1353–1358 (1986).
[13] J. M. Sabatier, H. E. Bass, L. N. Bolen, and K. Attenborough, ''Acoustically induced seismic waves,'' J. Acoust. Soc. Am. **80**, 646–649 (1986).
[14] J. M. Sabatier, H. Hess, W. P. Arnott, K. Attenborough, M. J. M. Roemkens, and E. H. Grissinger, ''*In situ* measurements of soil physical properties by acoustical techniques,'' Soil Sci. Soc. Am. J. **54**, 658–672 (1990).
[15] C. J. Hickey and J. M. Sabatier, ''Measurements of two types of dilatational waves in an air-filled unconsolidated sand,'' J. Acoust. Soc. Am. **102**, 128–136 (1997).
[16] W. P. Arnott and J. Sabatier, ''Laser-Doppler vibrometer measurements of acoustic to seismic coupling,'' Appl. Acoust. **30**, 279–291 (1990).
[17] E. M. Rosen, K. D. Sherbondy, and J. M. Sabatier, ''Performance Assessment of a Blind Test using the University of Mississippi's Acoustic/Seismic Laser Doppler Vibrometer (LDV) Mine Detection Apparatus at A. P. Hill,'' in Proc. SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets V, edited by A. C. Dubey *et al.*, paper 4037-74 (2000), pp. 656–666.
[18] C. B. Scruby and L. E. Drain, *Laser Ultrasonics Techniques and Applications* (Hilger, Bristol, Philadelphia and New York, 1990).
[19] K. Attenborough, H. E. Bass, and J. M. Sabatier, ''Elastic properties of the earth's surface,'' in *Handbook of Elastic Properties of Solids, Liquids, and Gases*, Vol. III, edited by Levy, H. E. Bass, and Stern (Academic, 2001).
[20] ''MATLAB User's Guide: Image Processing Toolbox,'' The Math Works Inc.
[21] D. Costley, J. Sabatier, and N. Xiang, ''Noise and vibration control for land mine detection,'' J. Acoust. Soc. Am. **107**, 2873 (2000).
[22] S. Yu, A. Gandhe, T. R. Witten, and R. K. Mehra, ''Physically based method for automatic mine detection using acoustic data,'' Proc. SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets V, edited by A. C. Dubey *et al.* (in press).
[23] R. Waxler, D. Velea, and J. M. Sabatier, ''A fast effective model for the scattering of normally incident sound off of a buried landmine,'' J. Acoust. Soc. Am. (submitted).
[24] N. Xiang and J. Sabatier, ''Fast *m*-sequence transform for laser-Doppler based mine detection,'' Proc. SPIE Conference on Signal Processing, Sensor Fusion, and Target Recognition VIII, edited by I. Kadar, SPIE 2720 (1999), pp. 390–396.
[25] P. Goggans, C. R. Smith, and N. Xiang, ''Increasing Speckle Noise Immunity in LDV-Based Acoustic Mine Detection,'' in Proc. SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets V, edited by A. C. Dubey *et al.*, paper 4038-80 (2000), pp. 719–724.
[26] N. Xiang and J. M. Sabatier, ''Iterative model-based automatic target recognition for acoustic landmine detection,'' J. Acoust. Soc. Am. **110**, 2740 (2001).

Beam intensity striations and applications

T. C. Yang[a)]
Naval Research Laboratory, Washington, DC 20375

(Received 28 February 2002; revised 7 November 2002; accepted 7 November 2002)

The single-element spectrogram for a continuous broadband signal, plotted as a function of range, has been shown to exhibit striated bands of intensity maxima and minima. The slope of the striations is an invariant of the modal interference and is described by a waveguide invariant parameter ''beta.'' The striation pattern is analyzed and modeled in this paper for the beam outputs of a horizontal line array obtained by conventional beamforming. Array beamforming makes it possible to measure the waveguide invariant parameter for weak signals due to the enhancement of signal levels by the array gain over that of a single element. It is shown that the signal beam spectrogram as a function of range exhibits the same striation pattern as that (predicted) for a single element. Specifically, for a broadside signal, the beam striation is identical to that of a single-element plus a constant signal gain. For a nonbroadside target, the signal beam intensity will be modified by a frequency-bearing dependent signal gain due to the signal spread over multiple beams, nevertheless the beam spectrogram retains the same striation pattern (slope) as for a single element. The sidelobe beams (outside the canonical cones containing the signal arrivals) exhibit an entirely different striation pattern as a function of frequency and range. For array processing, it is shown that a fast range-rate, close range target and a distant, slow range-rate interference source will have a different striation pattern (slope) in the corresponding beam spectrograms as a function of time, assuming no prior knowledge of the source ranges. The difference in the striations between the beam spectrograms can be used in array processing to suppress the interference contribution. A 5–7 dB interference suppression is demonstrated using simulated data. [DOI: 10.1121/1.1534604]

PACS numbers: 43.30.Bp, 43.30.Wi, 43.60.Gk [WMS]

I. INTRODUCTION

Modal interferences are evident in acoustic signals propagated over distances.[1] For example, constructive or destructive modal interferences are responsible for the peak and valley of narrowband transmission loss (TL) versus range. Likewise, matched field processing requires a replica field consisting of coherent sums of normal modes where the modal interference is critical for source localization. For a coherent broadband signal, the signal energy travels via the modal group velocities. The deep and Arctic Oceans belong to one group in which the low order modes travel with lower group velocities. Consequently, the low order modes arrive last. The shallow water waveguide generally exhibits an opposite behavior in that the low order modes travel faster and arrive earlier than the high order modes. Modal arrivals and dispersions can be studied using the time-frequency analysis as has been done for explosive signals.[1] The slope of the modal intensity as a function of time and frequency is proportional to the source range.[2]

Individual mode arrivals usually cannot be identified in arrivals of continuous broadband signals, which often do not have a well-defined wave form. Although the frequency components are in this case incoherent with respect to each other, the broadband intensity spectrum plotted as a function range retains some features of modal interference. The single-element spectrogram has shown striated bands of intensity maxima and minima as a function of range, corresponding to constructive and destructive interference of normal modes. The slope of the striations has been claimed to be invariant to the details of the waveguide and can be characterized by a scalar parameter, β, referred to as the waveguide invariant.[3–5] It was shown that for a constant sound speed profile (SSP) or that with a linearly increasing square-of-refraction-index, the slope of the striation pattern is independent of the mode number, the acoustic frequency, the source-receiver depth.[3–5] Extension of the waveguide invariant to a range dependent environment,[6,7] and applications to time-reversal[8] and source localization via matched-field processing sidelobes[9] can be found in recent publications. One notes that for a general SSP environment, the slope of the intensity striations can vary with respect to the mode number, the (center) acoustic frequency, and the source-receiver depth. Although β is in these cases no longer a ''true invariant'' in the usual sense, it will still be referred to as the invariant parameter in this paper.

We study in this paper the striation pattern of a beam intensity spectrogram as a function of frequency and range, obtained by conventional beamforming of a horizontal array. It is noted that the study of the intensity striation requires a high signal-to-noise ratio (SNR) lest the striation be masked by the noise. Array beamforming provides a high array gain and is used to enhance the SNR. As a result, the modal interferences, or striations, are better identified in the beam intensity than the single element intensity spectrogram.

a)Electronic mail: yang@wave.nrl.navy.mil

We shall demonstrate.

(1) The characteristic features of beam striations, which are beam angle dependent. At the signal arrival angles, the beam spectrogram as a function of range exhibits the same striation pattern (and the same slope) as that of a single element. At off-signal directions, the beam spectrogram exhibits a different striation pattern than the signal beam.

(2) A simple two-dimensional fast Fourier transform (FFT) to measure the striation slope. The slope of the FFT spectrum is analytically related to the waveguide invariant parameter, β.

(3) A simple algorithm based on beam spectrogram striations as a function of time to suppress the interference contribution in array processing. It is noted that the signal is often masked by a strong interference source in the single-element spectrogram. In the beam domain, the signal and interference often show up in different beams. The difference in striations between the signal and interference beams can be used to suppress the interference in array beamforming. This method of interference suppression is complimentary to other signal processing approaches, such as interference nulling.

This paper is organized as follows. In Sec. II, we briefly review the waveguide invariant theory and introduce a two-dimensional Fourier transform, which yields a simple measurement of the waveguide invariant parameter from the spectrum slope. In Sec. III, we calculate the beam intensity striation and demonstrate that the beam striation varies with the beam angle. We show that the beam spectrogram at the signal arrival angles exhibit the same striation pattern and slope as a function of range as that of a single element. At off-signal directions, the beam spectrum exhibits a different striation pattern. In Sec. IV we present array processing incorporating beam striations. Source ranges are assumed unknown and the beam spectrogram data are now a function of time. We consider a close range, fast range rate (e.g., fast approaching) submerged source in the presence of a distant, slow range rate (e.g., slow moving) but loud surface ship (interference). Using simulated data, we demonstrate a 5–7 dB interference rejection based on the difference in the beam striations. The simulations were done for the SWellEX-96 environment,[10] for which the modeled striation patterns have been compared favorably with the measurements at an element level.[6] Section V summarizes the paper.

II. INTENSITY STRIATIONS: SINGLE ELEMENT

The derivation of the waveguide invariant β, including Eqs. (1)–(7), follows closely Ref. 3. The pressure field for a receiver at range r and depth z can be expressed in terms of the normal modes as

$$p(\omega)=\sum_m \sqrt{2\pi}\exp(-jk_m r)\psi_m(z)\psi_m(z_s)/\sqrt{k_m r_j} \equiv \sum_m A_m \exp(-jk_m r), \tag{1}$$

where the mode amplitude A_m and mode wave number k_m are functions of the acoustic frequency ω, and $\psi_m(z)$ is the mode depth function. The intensity is given by

$$I(\omega)=|p(\omega)|^2 \equiv \sum_m |A_m|^2 + \sum_{m\neq n} A_m^* A_n \exp(j(k_m-k_n)r). \tag{2}$$

The intensity maximum is determined by the following condition, ignoring the slow variation of A_m with respect to frequency and range:[11]

$$\delta I \approx j\sum_{m\neq n} A_m^* A_n e^{j(k_m-k_n)r}\delta((k_m-k_r)r)=0. \tag{3}$$

The stationary phase condition can be expressed as

$$\delta((k_m-k_n)r)=\left(\frac{dk_m}{d\omega}-\frac{dk_n}{d\omega}\right) r\, d\omega+(k_m-k_n)dr=0. \tag{4}$$

One finds

$$\frac{d\omega/\omega}{dr/r}=-\left(\frac{k_m-k_n}{\omega}\right)\Big/\left(\frac{dk_m}{d\omega}-\frac{dk_n}{d\omega}\right) =-\left(\frac{1}{\nu_m}-\frac{1}{\nu_n}\right)\Big/\left(\frac{1}{u_m}-\frac{1}{u_n}\right)\equiv\beta_{mn}, \tag{5}$$

where $\nu_m=\omega/k_m$ is the phase velocity and $u_m=d\omega/dk_m$ is the group velocity of the mth mode. For a waveguide of constant sound speed profile overlying a hard bottom (the Pekeris waveguide), β_{mn} is the same for all the modes,[3]

$$\beta_{mn}=\beta=1$$

and for a waveguide with the square of the index of refraction $n^2(z)=c_0^2/c^2(z)$ increasing linearly with depth,

$$\beta_{mn}=\beta=-3.$$

For these waveguides, β is called an invariant, since the spectral slope as expressed by the left-hand side of Eq. (5) will be the same for all frequencies and for all source and receiver depths.

Figure 1 shows, as an example, the intensity spectrum for a Pekeris waveguide covering a frequency range from 150 to 200 Hz and a source–receiver range from 2800 to 4400 m. The water column has a sound speed of 1500 m/s, with a bottom at depth of 213 m. (The water depth, range, and frequencies are chosen to coincide with the SwellEx96 simulation in the following.) The source and receiver are at the middle of the water column. One should note that although the pattern of intensity striation should be independent of the source–receiver depth and/or the water depth, the actual intensity level will vary depending on the frequency and source–receiver depths.

An intuitive interpretation of Eq. (5) can be given based on Fig. 1. Note that at a fixed frequency, the acoustic intensity as a function of range is proportional to TL. The difference in range, δr, between two adjacent intensity maxima is the well-known modal cycle distance

$$\delta r=\frac{2\pi}{k_m-k_n}=2\pi\left[\omega\left(\frac{1}{\nu_m}-\frac{1}{\nu_n}\right)\right]^{-1}. \tag{6}$$

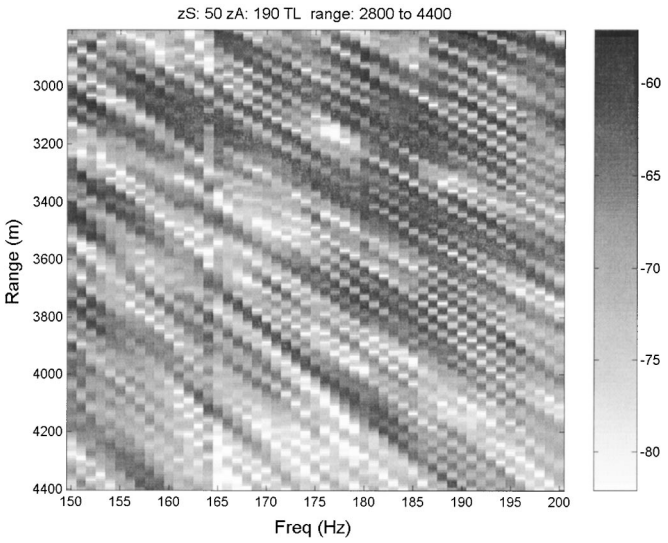


FIG. 1. Intensity as a function of frequency and range for a single receiver in a Pekeris waveguide. There are 200 range grids and 50 frequency bins.

This expression can be derived from the expression for the intensity maximum corresponding to a constructive interference between two modes, given by $k_m r = k_n r + l2\pi$, where l is an integer. Likewise, one finds that at a fixed range, the intensity maximum is obtained when the two modes arrive in phase: $T_m \omega = T_n \omega + l2\pi$. Hence, the difference in frequency, $\delta\omega$, between two adjacent intensity maxima is given by

$$\delta\omega = \frac{2\pi}{T_m - T_n} = 2\pi\left[r\left(\frac{1}{u_m} - \frac{1}{u_n}\right)\right]^{-1}, \tag{7}$$

where we used the fact that the mode intensity travels with its group velocity. The ratio of Eqs. (6) and (7) yields Eq. (5).

To determine the waveguide invariant parameter β, we take a two-dimensional (2D) Fourier transform of Fig. 1 and the result is shown in Fig. 2. The slope of the mainlobe in Fig. 2 is related to the invariant parameter β. To see this, we

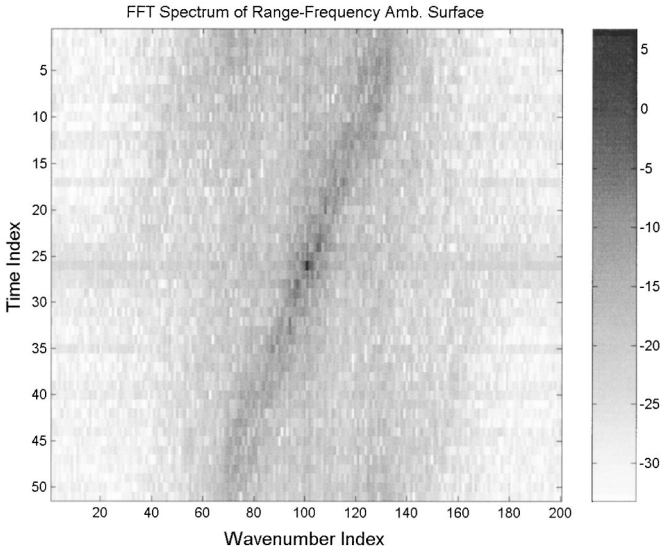


FIG. 2. Two-dimensional FFT spectrum of the intensity surface as shown in Fig. 1. The x and y axes are the Fourier transform variables conjugate to range and frequency.

use a discrete representation of the intensity spectrum. Ignoring the first term in Eq. (2) which yields a dc term in the FFT results, and expanding the second term in a Taylor series yield

$$I(l,h) = \sum_{m \neq n} A_m^* A_n \exp(j(\hat{k}_m - \hat{k}_n) r_0) \exp\left[j(\hat{k}_m - \hat{k}_n) l\Delta r + j\left(\frac{d\hat{k}_m}{d\omega} - \frac{d\hat{k}_n}{d\omega}\right)(h\Delta\omega) r_0 \right], \tag{8}$$

where $\hat{k}_m \equiv k_m(\omega_0)$, and l, h are the range and frequency indices, respectively. The range and frequency coordinates are given by

$$r_l = r_0 + l\Delta r, \quad l = -L/2,...,0,1,..,L/2,$$

$$\omega_h = \omega_0 + h\Delta\omega, \quad h = -H/2,...0,1,...,H/2,$$

where Δr and $\Delta\omega$ are the range and frequency bin width. Taking a discrete Fourier transform of $I(l,h)$,

$$F(f,g) = \sum_{l,k} \exp(-j2\pi[gl + fh]) I(l,h), \tag{9}$$

where f and g are the FFT variables conjugate to frequency and range, respectively, one finds a trajectory of points in the (f, g) plane associated with a pair of modes m and n, with a slope S_{mn} given by

$$S_{mn} = \beta_{mn} \frac{\omega_0 \Delta R}{r_0 \Delta\Omega}, \tag{10}$$

where $\Delta R \equiv L\Delta r$ and $\Delta\Omega \equiv H\Delta\omega$ are the total increment of range and frequency of the intensity spectrum (e.g., Fig. 1). The FFT approach shows clearly the intensity striation pattern as exhibited in Fig. 2, where f is denoted as (delay) time, and g is referred to as a wave number variable. The line slope yields S_{mn}.

In a general sound speed profile environment, the slope of the striation pattern can cover a range of values, yielding a distribution of β. The waterborne modes and the boundary interacting modes normally have a different striation slope (e.g., $\beta = -3$ and $\beta = 1$, respectively). The acoustic field, consisting of both the waterborne and boundary interacting modes, thus has a striation slope in between -3 and 1. The distribution depends on the source and receiver depth and the acoustic frequency. The FFT approach will show the respective slopes of the different modes.

Figure 2 shows a strong peak at the center of the plot. The origin of this peak is the first term of Eq. (2), which produces a dc component in the FFT spectrum as remarked previously before Eq. (8).

III. BEAM INTENSITY STRIATIONS

In this section, we study the beam intensity striations, obtained by conventional beamforming of a horizontal line array (HLA). Again, we employ a normal mode representation of the pressure field

$$p_j(\omega)=\sum_m \sqrt{2\pi}\exp(-jk_m r_j)\psi_m(z)\psi_m(z_s)/\sqrt{k_m r_j}$$

$$\equiv\sum_m A_m\exp(-jk_m r_j), \tag{11}$$

where r_j is the range from the source to the jth element, and all elements are at the same depth z. Assuming an array of N uniformly spaced elements, with spacing d, the conventional beam outputs are obtained by

$$B(\theta,\omega)=\left|\sum_{n=1}^{N} e^{j(n-1)k_0 d\sin\theta}p_n(\omega)\right|^2, \tag{12}$$

where the angle θ is measured from the broadside of the HLA and $k_0=\omega/c_0$ with c_0 being the sound speed at the HLA depth. To carry out the analytical calculations, we shall take a far field expansion of the source–receiver range[11]

$$r_n\cong r_0-(n-1)d\sin\theta_T, \tag{13}$$

where r_0 is the range to the first element of the array and θ_T is the signal (or target) bearing. We then find that

$$B(\theta,\omega)=\left|\sum_m A_m\exp(-jk_m r_0)\sum_{n=1}^{N}\exp(j(n-1)\times d[k_m\sin\theta_T-k_0\sin\theta])\right|^2$$

$$=\left|\sum_m A_m\exp(-jk_m r_C)\times\mathrm{sinb}(k_m\sin\theta_T-k_0\sin\theta)\right|^2, \tag{14}$$

where

$$r_C\equiv r_0-\frac{(N-1)d}{2}\sin\theta_T$$

is the range to the source from the center of the array, and

$$\mathrm{sinb}(X)\equiv\sin\left(\frac{Nd}{2}X\right)\Big/\sin\left(\frac{d}{2}X\right). \tag{15}$$

We note that the square of the above-given function, with $X=k_0(\sin\theta_T-\sin\theta)$, is the usual conventional beam pattern of a HLA with N elements, steered to the signal direction θ.

Returning to Eq. (14), let $X_m=k_m\sin\theta_T-k_0\sin\theta$, the beam intensity spectrum can be written as

$$B(\theta,\omega)=\sum_m |A_m\,\mathrm{sinb}(X_m)|^2+\sum_{m\neq n}A_m^* A_n\times\exp(-j(k_m-k_n)r_C)\mathrm{sinb}(X_m)\mathrm{sinb}(X_n). \tag{16}$$

We note the following.

(1) For a plane wave arrival, Eq. (16) reduces to the conventional beam pattern. A plane wave is equivalent to a single mode arrival with the mode phase velocity equal to the sound speed, $k_m=k_0$. One then finds $B(\theta,\omega)\approx\mathrm{sinb}^2(X)$, where $X\equiv k_0(\sin\theta_T-\sin\theta)$, which is the plane wave beam pattern function.

(2) For arrivals consisting of many modes, one notes then that the first term of Eq. (16) is a summation of modal beam patterns. The second term of Eq. (16) is the ''modal beam interference'' term.

(3) When the signal is at the broadside direction, $\mathrm{sinb}(X_m)=N$ at the signal-look direction ($\theta=\theta_T=0°$). The beam intensity is N^2 times that of a single element.

(4) For a nonbroadside signal, $X_m=(k_m-k_0)\sin\theta_T$ at the signal-look direction ($\theta=\theta_T\neq 0°$). We find that $X_m d$ is a small number in many environments since the wave number differences between the modes are usually small (k_0 equals approximately the mode wave number of the lowest contributing mode). In that case, the function $\mathrm{sinb}(X_m)$ is slowing varying with respect to frequency. At the signal arrival angle we will show in the following that $\mathrm{sinb}(X_m)$, is a constant having a value close to N.

(5) While the function $\mathrm{sinb}(X_m)$ may have different values, the important thing is that it is not a function of range. Hence, although it may modify the beam intensity level, it is not expected to modify the striation slope.

(6) The beam intensity spectrogram has in fact the same striation slope as the element spectrogram. To see this, note that the mode interference term in the beam spectrum, Eq. (16), is identical to the mode interference term in the single-element spectrum, Eq. (2), except for the constant $\mathrm{sinb}(X_m)\mathrm{sinb}(X_n)$ factor. Applying the same FFT analysis (of Sec. II) to Eq. (16), the beam FFT spectrum yields the same slope S_{mn} as given by Eq. (9), since the constant sinb terms cancel out in the ratio to get the slope.

The above-noted observations will be verified using simulated data. Note that a horizontal line array has a unique feature that the endfire beam is different from the broadside beam. The modal arrivals for a broadside source show up all in one beam. The signal for a near endfire source is split into many beams, the beams containing the dominant signal energy may be offset from the signal-look direction by many degrees. Hence, we need to distinguish the beams at the signal-arrival angles from the beam at the signal-look direction. For a near endfire source, the mth mode of the signal should arrive at a beam angle determined by

$$\sin\theta_m=\sin\theta_T\cos\phi_m=k_m\sin\theta_T/k_0, \tag{17}$$

where ϕ_m is the vertical arrival angle of the mth mode. The signal beams are no longer at the signal-look direction. At the signal-arrival beam associated with the mth mode, i.e., $\theta=\theta_m$, one finds $\mathrm{sinb}(X_m)=N$. In the neighborhood of θ_m, $\mathrm{sinb}(X_m)\sim N$. We thus find that the signal-arrival beams exhibit a striation pattern (as a function of frequency and rage) closely parallel to that of a single element, with a beam-angle dependent modal signal gain given by Eq. (16). This will also be illustrated in the following by the simulated data.

For numerical simulations of the beam striation pattern, we shall use the SWellEX-96 acoustic environment[10] where the intensity striation pattern of a single element has been previously modeled and compared with the measurements for the lowermost hydrophone on a vertical line array.[6] For

this simulation, we shall consider a bottom mounted HLA similar to that deployed during this experiment. Based on the above-mentioned theoretical analysis, we have reasons to believe the modeled beam striation pattern for the HLA should also agree with the measurements, but this has not been verified.

The SwellEX-96 has a downward refractive sound speed profile. The water column has a depth of 213 m, with a near constant sound speed of 1488 m/s from 100 m to 213 m. The sound speed increases slightly to 1494 m/s as depth decreases to 35 m. From 35 m to the surface, the sound speed increases more rapidly to 1522 m/s at the surface. The sound speed profile is plotted in Fig. 1 of Ref. 10. For this profile, β is expected to be close to one.[6] Since the purpose of this paper is not to model the value of β as a function of the acoustic environment, the details of the sound speed profile are not critical. In fact, we assume no prior knowledge for the value of β.

For the bottom mounted HLA, we use an array of a similar aperture as that used in the SwellEX-96 experiment. During the SWellEx-96, a HLA (i.e., the AODS-N) was deployed in the proximity of the vertical array (see Ref. 10, Fig. 1). The AODS-N array had an aperture of ~240 m and contained 31 (working) elements which are unevenly spaced.[10,12] For the simulation, we shall assume a HLA of approximately the same aperture and the same number of phones, but for simplicity we shall assume that the phones are evenly spaced. (This particular array has a design frequency $f_0 \approx 95$ Hz.) The source is deployed at a depth of 50 m. We shall consider two cases: a broadside source and a near endfire source at a bearing of 79° from the broadside, the latter corresponding to event E in Ref. 6. For the continuous broadband signals, we shall consider a bandwidth of approximately half of the design frequency of the array. The pressure field is calculated using the KRAKENC normal mode code.[13]

We will show a single-element spectrogram plus several beam spectrograms as a function of range to compare the intensity striation patterns. The beam spectrograms are for a signal beam at 0° (broadside), 79°, 63°, and a sidelobe beam 45° from the signal direction. Figure 3 shows the intensity striation pattern for a single-element on the HLA for a source covering a range of 2800–4400 m as described earlier. Figure 4 shows the beam intensity striation pattern for a broadside source covering the same range. For this case, the beam striation is (nearly) identical to that of Fig. 3 (as suggested by the above-given analysis) with an increase in the intensity level due to array signal gain of ~30 dB.

For comparison, Fig. 5 shows the beam striation pattern for a near endfire source (i.e., $\theta_T = 79°$ as in event E) steered at the signal look direction, $\theta = \theta_T = 79°$ [Fig. 5(a)] and signal arrival direction, $\theta = 63°$ [Fig. 5(b)]. The beam angles for the signal arrival beams are predicted using Eq. (17). Using the modal wave numbers calculated at the center of the frequency band, one finds that the first, eighth, and fifteenth modes arrive at a beam angle of approximately 76°, 64°, and 50°, respectively. Because of the shallow source depth (at 55 m), the lowest order (first to third) modes are weakly excited. As a result, the signal energy appears in higher order modes

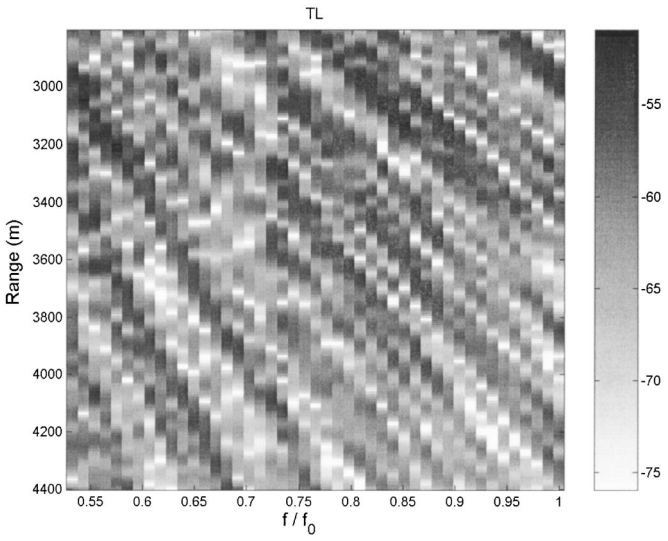


FIG. 3. Intensity as a function of frequency and range for a bottom receiver in the SWellEX-96 waveguide. The frequency is displayed as a ratio to the design frequency of the array f_0. There are 200 range grids and 45 frequency bins.

covering a beam angle from 60° to 70°. The signal beam response is further discussed in Sec. IV. Figure 5(b) shows the beam intensity as a function of range using the beam at 63° as an example.

One notes that the signal-look beam in Fig. 5(a) has a lower intensity level than the signal-look beam at broadside, but it has the same striation slope as in Fig. 4. The lower intensity level in Fig. 5(a) is typical of the signal-look beam for a near endfire source when the signal arrival angle is offset from the signal-look direction (as remarked earlier). The beam intensity at 63° is about 4 dB below that of Fig. 4 because the signal energy is split in several beams (from 60° to 70°). It has the same striation slope as that of a broadside target (Fig. 4) in accordance with the above-given analysis.

Next we show the beam striation patterns at a nonsignal

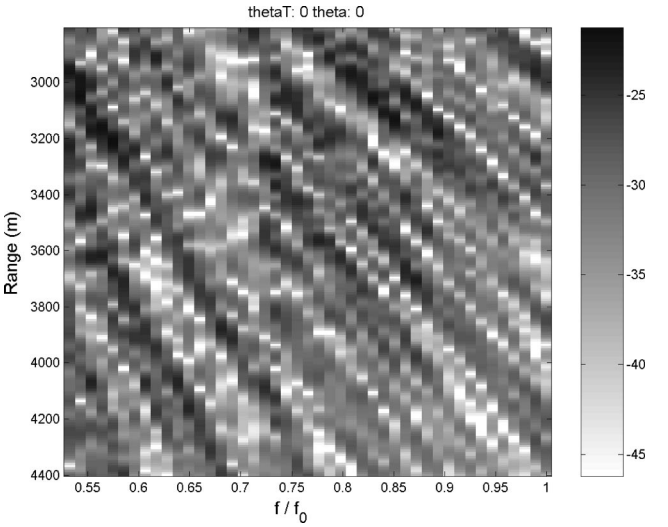
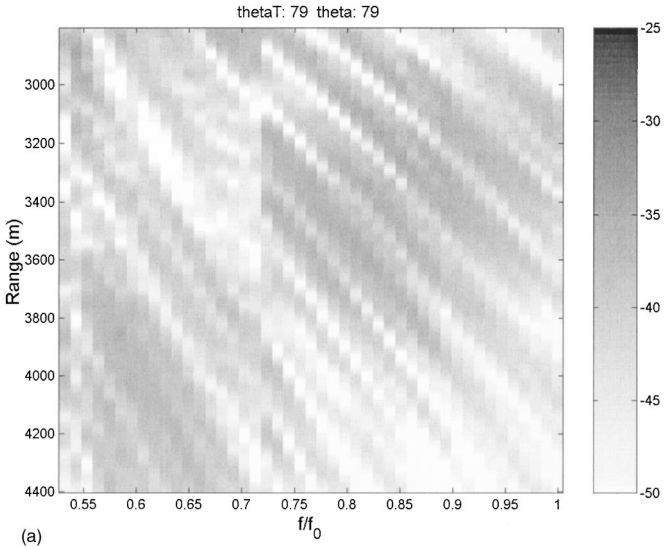


FIG. 4. Beam intensity as a function of frequency and range for a bottom HLA in the SWellEX-96 waveguide. The beam is steered at the target, located at the broadside direction. Same frequency and range as in Fig. 3 caption.



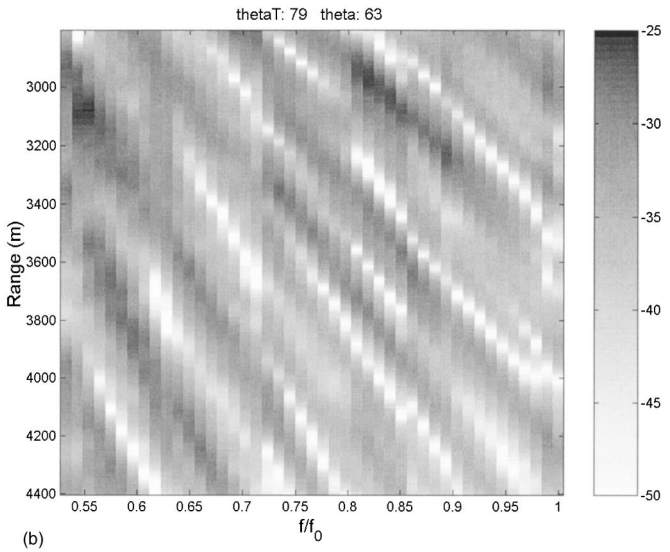


FIG. 5. Same as Fig. 4 except that the target bearing is 79° from the broadside direction. The beam is steered at the signal look direction in (a) and at a signal arrival direction in (b). Same frequency and range as in Fig. 3 caption.

look direction, the so-called sidelobe beam. Consider a sidelobe beam separated by 45° from the signal bearing. Figure 6 shows the intensity striation of a sidelobe beam at 34° for a signal at 79° from the broadside. We note that this sidelobe beam is dominated by a ''vertical'' striation pattern. This is typical of many sidelobe beams. The reason can be seen as follows. Note that Eq. (16) has two terms. The second term, the modal interference term, carries a range-dependent phase, which is responsible for the intensity striations as a function of range. The first term of Eq. (16) is a summation of the modal beam patterns (see the earlier discussions). It has a weak range dependence due to the decreasing mode amplitudes with range. Thus it produces either no striation or a ''vertical'' striation pattern; the apparent ''vertical'' striation pattern is due to the source spectrum. Note that at the signal arrival angle, $\sin b(X_m)$ is near unity, as discussed earlier. The contribution of the second term is as large as the first term. For the sidelobe beams, $\sin b(X_m)$ is zero or a small number, its value being frequency dependent. The

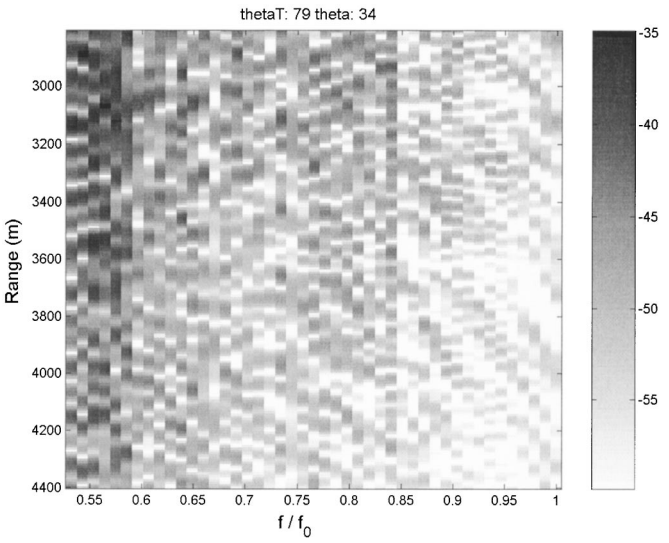


FIG. 6. Beam intensity for a sidelobe beam at 34°, for a target at 79° from the broadside direction. Same frequency and range as in Fig. 3 caption.

modal interference terms tend to cancel each other out when integrated over frequencies, due to the rapid change of $\sin b(X_m)$ with frequency at the sidelobe directions. In contrast, the modal beam patterns' terms are positive definite and added up constructively. The modal beam patterns' terms dominate at the sidelobe direction and produce an almost range-independent, vertical striation pattern.

The summary story is that the signal-arrival beams exhibit the same intensity striations as a function of frequency and range as for a single-element (as predicted by the waveguide invariant theory). The beam intensity distribution is identical to that of a single element, enhanced by the array signal gain. At broadside, the signal gain is near theoretical. At near endfire directions, the signal may be spread over 3 several beams yielding a lower (angle dependent) signal gain. Although the signal beam intensity may vary depending on the signal bearing, the striation slope remains the same as that for a single element. For the sidelobe beams (beams not associated with multipath arrivals of the signal) we find that they do not exhibit the striations associated with the waveguide invariants. They are dominated by vertical striations associated with the modal beam patterns.

IV. APPLICATIONS TO ARRAY PROCESSING

In this section, we use the beam striation pattern to identify a fast range-rate source (i.e., a source moving rapidly in range) against a slow range-rate distant interference (surface ship), and we demonstrate interference suppression based on the signal striation. The beam striation pattern in this section refers to the intensity striation in the beam spectrogram as a function of time, as data are given as a function of time. The source ranges are assumed unknown to the processor. No prior knowledge of the striation parameter, β, is assumed, nor is β assumed an invariant. Although the striation parameter could be measured directly from the data (Sec. III), the measurement is not required for array processing.

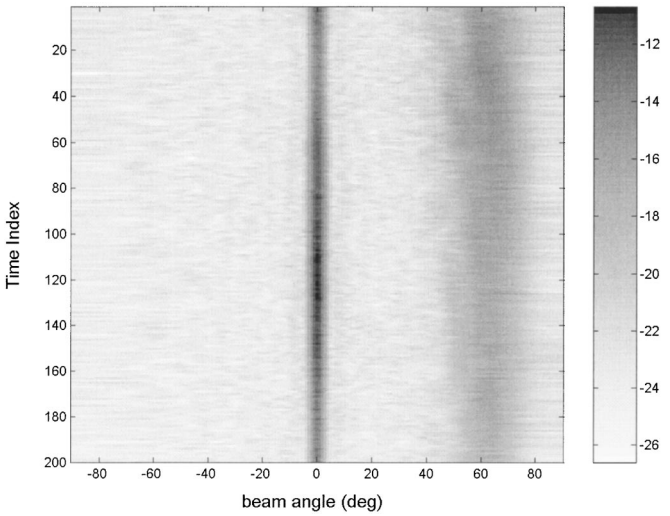


FIG. 7. Conventional beam time record for a target at 79° from the broadside direction with an input SNR of −5 dB. A loud surface ship is found at broadside direction.

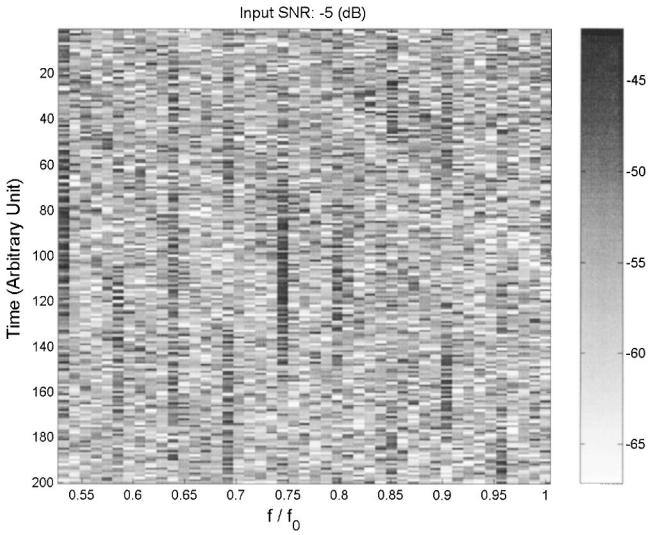


FIG. 8. Intensity as a function of frequency and time for a bottom receiver. The difference between this figure and Fig. 3 is the presence of noise and the interference source. There are 200 time grids and 45 frequency bins.

The proposed array processor takes advantage of the differences in the striation patterns between a signal and an interference source. This difference is clearly illustrated using the FFT spectrum. We use FFT for its simplicity, speed, and the fact that the spectrum slope is analytically related to the β parameter. Our purpose is to illustrate potential array performance improvement derived from striation processing. More sophisticated algorithms can be developed in the future using other methods, such as the Radon or Hough transform.

The data are simulated using the SWellEX-96 environment as described earlier. We consider two scenarios. For the first scenario, the source is located at a near endfire bearing of 79° as before covering a range of 2800–4400 m in a time period of 200 units. (The range is not known to the array processor.) We consider a low SNR case by including a random noise that is temporally Gaussian distributed and spatially uncorrelated between the array elements. The noise is 5 dB above the received signal level at the input of the receiver. Next, we include a loud distant surface ship located at the broadside direction. The ship is located at a range of 20 km from the array with a radial velocity one-quarter of the source velocity. The ship sound radiation depth is assumed at 5 m. The ship projects continuous broadband signals with a source level 20 dB higher than the signal source, and many narrowband signals that are (an additional) 14 dB higher than the broadband signals. Again, the ship range and speed are unknown to the array processor. The second scenario is identical to the first one except that the source is located at the same bearing as the interference (i.e., broadside of the array). The purpose of the first scenario simulation is to demonstrate interference suppression when the interferer is separated in bearing from the target source. The purpose of the second scenario simulation is to demonstrate target detectability using the beam striation when both the source and interferer are located at the same bearing. Note that conventional array processing cannot detect the presence of a target for this (second) case since the target signal is masked by the strong interference field.

For simplicity, we shall assume that both the source and interferer are moving in a radial direction, keeping their bearing unchanged. In practice, they are likely to change bearing as well as range at the same time. In that case, one will replace the beam spectrogram by the so-called scissorgram, which is the beam spectrogram following the source bearing. The analysis remains the same.

Figure 7 shows the conventional beam time record, where the conventional beam outputs are integrated over the signal frequencies. It is apparent that the interference field is significantly (4–5 dB) stronger than the signal. Note that for the near endfire source at a bearing of 79°, the signal arrives at the HLA predominantly at beam angles from 60° to 70°.

Figure 8 shows the intensity spectrogram as a function of time for a single element on the HLA. The single-sensor result is included to show that the interfering source drowns out the desired signal. One notes that the signal striation as seen in Fig. 3 is hardly visible in Fig. 8. The striation pattern of Fig. 8 is dominated by the narrowband signals from the surface ship as well as the background noise. (Narrowband signals show up as straight lines in the beam spectrogram.) This is confirmed by the 2D FFT spectrum shown in Fig. 9, which exhibits a vertical line instead of a sloped line as indicated by Fig. 2. The poor results at the sensor level are the motivations for studying the striations in the beam space, which has a capability to discriminate the signal against the interfering field and noise. To demonstrate noise discrimination, we shall examine the 2D FFT spectrum of a random noise noting that noise is 5 dB above the signal at the sensor level.

Figure 10 shows intensity distribution as a function of frequency and time for a Gaussian random noise. Figure 11 shows the FFT spectrum of the time-frequency surface of Fig. 10. The FFT spectrum of the noise intensity distribution is dominated by a dc component at the origin as indicated by the sharp peak at the center of the figure. The reason is that random noise is incoherent. It is independent of range and shows no striation with respect to time. For an ocean wave-

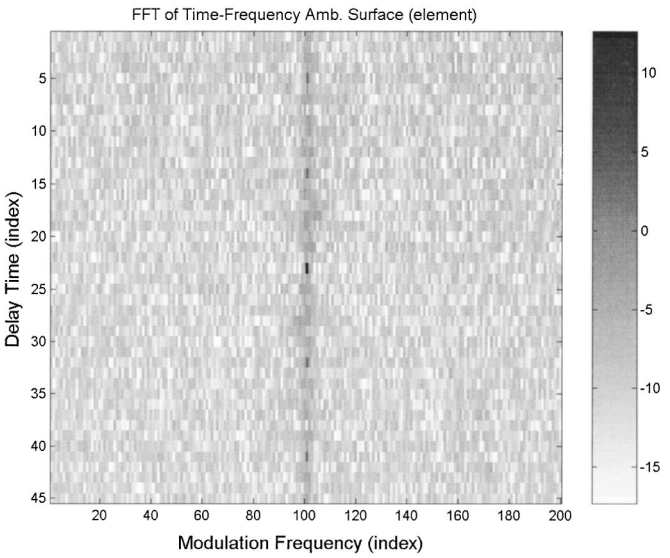


FIG. 9. Two-dimensional FFT spectrum of the intensity surface shown in Fig. 8. The x and y axes are the Fourier transform variables conjugate to time and frequency.

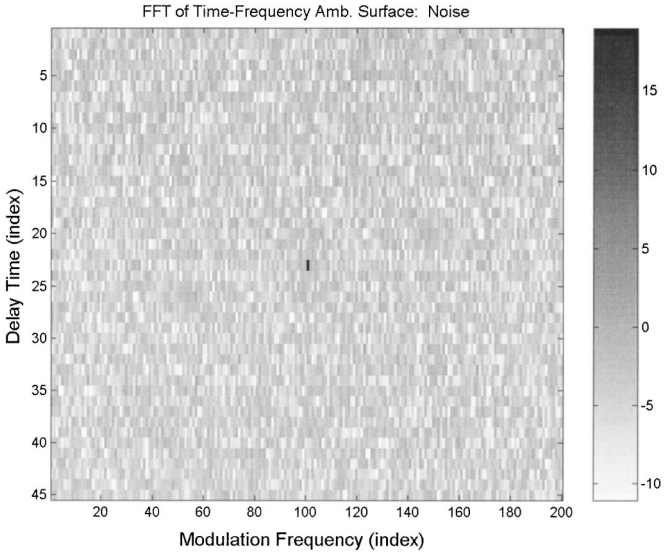


FIG. 11. Two-dimensional FFT spectrum of the noise intensity surface shown in Fig. 10. The x and y axes are the Fourier transform variables conjugate to time and frequency.

guide, the ambient noise from incoherent noise sources is also represented by an incoherent sum of normal modes[14] and contains no mode interference term, hence no striations. Note that the 2D FFT spectrum of the noise (Fig. 11) is significantly different from that of the signal (Fig. 2). One can employ spectral analysis to suppress the noise.

Let us examine the striation pattern of the signal beam for the first scenario case first. Figure 12 shows the beam intensity spectrogram for a beam steered at 63° from the broadside direction, representing one of the signal-arrival beams (based on Fig. 7). Contrasting Fig. 12 with Fig. 8 of a single element, one sees clear evidence of the signal striation (in the signal beam) despite the fact that the noise is 5 dB above the input signal level and the interfering ship is 20 dB louder than the source. Again the striation is verified by the FFT spectrum of the 2D intensity distribution (shown in Fig. 13), which resembles Fig. 2. We noted that beamforming has enhanced the signal against the noise, and has separated the signal from the interfering source, making the identification of the signal striation easier. We note that for a large aperture HLA, the sidelobe level is often 15–20 dB down (as is the case here) compared with the main lobe level. The leakage from the interference beam to the signal beam is thus small and can be treated as background noise. The FFT spectrum shows indeed that the loud interfering ship has a negligible impact on the signal beam striations.

Figure 14 shows, on the other end, the beam intensity striation pointing to the interfering surface ship. The beam intensity is dominated by the narrowband signals which show up as vertical lines in the beam spectrogram. The incoherent broadband signals have a smaller (more vertical) striation slope $(d\omega/dt)$, since the ship is located at a range five to seven greater than the signal source, and it travels

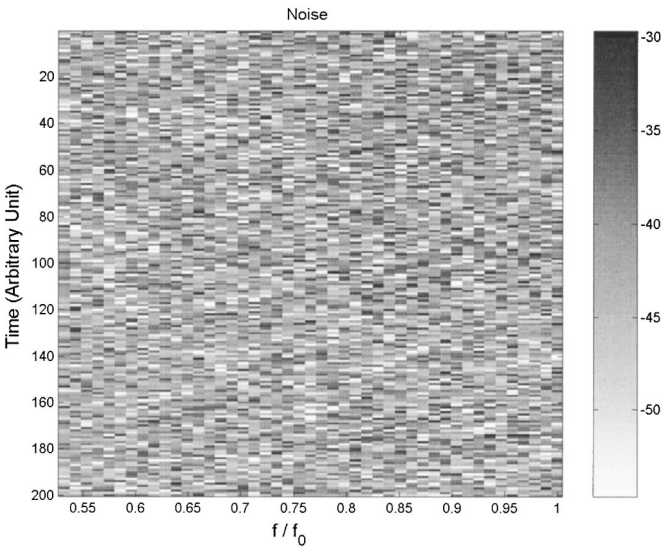


FIG. 10. Beam intensity for Gaussian white noise.

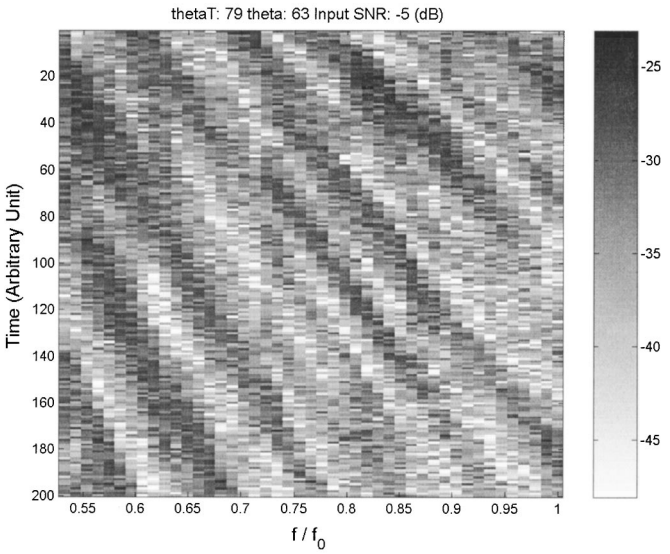


FIG. 12. Beam intensity as a function of frequency and time steered to the target, located at 79° from the broadside direction. Same frequency and time as in Fig. 8 caption.

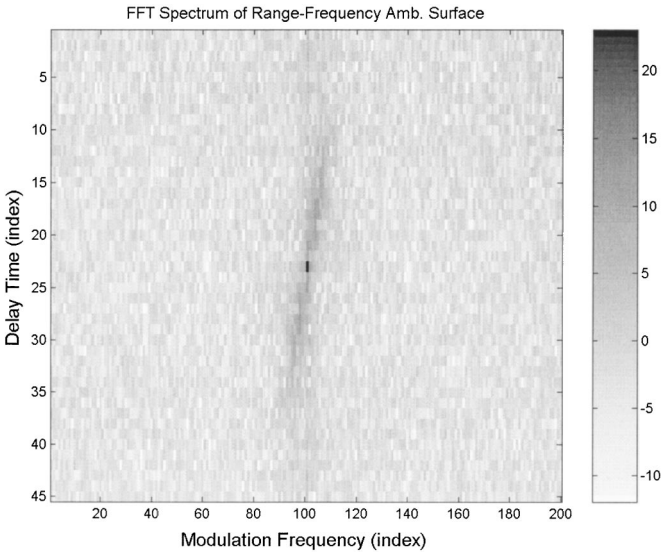


FIG. 13. Two-dimensional FFT spectrum of the beam intensity surface shown in Fig. 12. The x and y axes are the Fourier transform variables conjugate to time and frequency. Beam striation is indicated by the slope of the spectrum line.

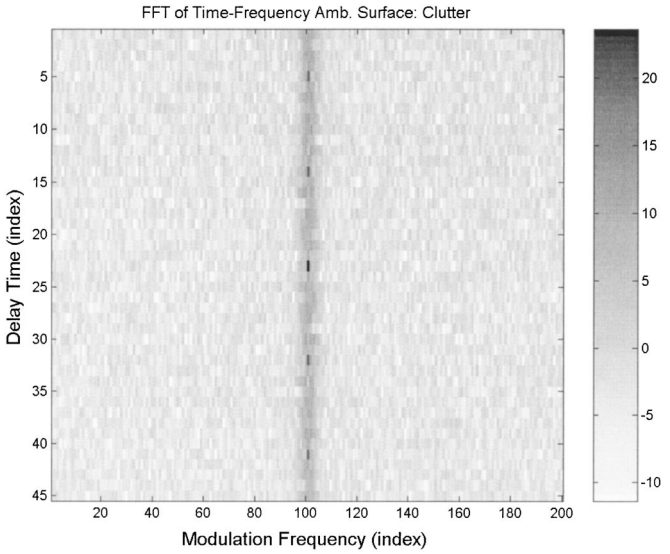


FIG. 15. Two-dimensional FFT spectrum of the beam intensity surface steered to the surface ship, as shown in Fig. 14. The x and y axes are the Fourier transform variables conjugate to time and frequency. The spectral line is practically vertical.

with a radial speed (and hence the range increment) one-quarter of the source speed. As a consequence, it has a much slower rate of variation with respect to frequency according to Eq. (5). The slow rate of frequency striation is evident in the FFT spectrum plot by its near vertical stripe pattern as displayed in Fig. 15. Comparing Fig. 8 with Fig. 14, one notes that the striation pattern for a single element is dominated by the contribution from the surface ship.

The above-mentioned simulations showed that signal striations are more evident in the array beam output focused at the signal arrival direction than at an element level. The FFT spectrum of the 2D intensity distribution shows clearly the differences of the striation patterns for a fast range-rate, close range source versus a long range, slow range-rate interferer. The difference in striations can be used as a cue for detection of a fast range-rate object, which will be addressed later.

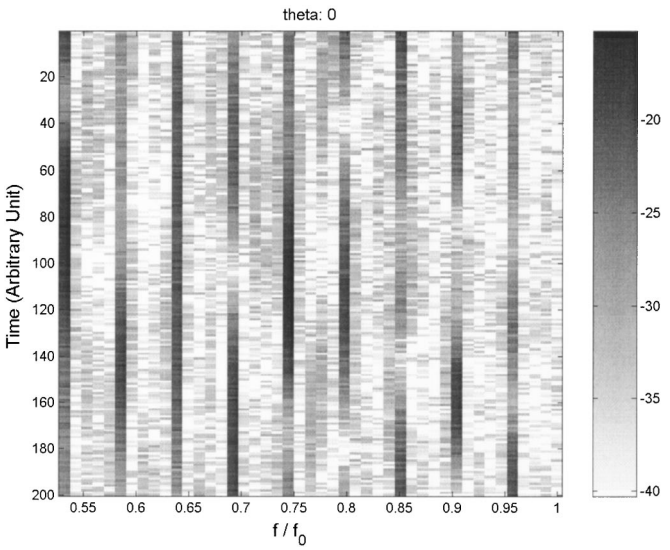


FIG. 14. Beam intensity as a function of frequency and time steered to the surface ship, located at the broadside direction. Same frequency and time as in Fig. 8 caption.

One can take advantage of the above-mentioned differences between beam striation patterns of the signal, surface interference, and noise to suppress the contribution of the interference field and noise thus resulting in an improved output SNR and signal-to-interference ratio over that of conventional beamforming. The basic objective of the signal processing algorithm will be to remove the FFT components associated with the vertical striation lines in the FFT spectrum of the 2D intensity surfaces. A simple algorithm is used to illustrate this approach, based on thresholding the FFT spectra of the beam intensity function. The threshold is set by the FFT spectrum of the beam associated with the interfering ship, which is assumed known since the interfering ship produces the highest beam response in the conventional beam output. Using Fig. 15 as an example, a threshold is set at -20 dB from the peak of the FFT spectrum surface. This identifies a filter grid consisting of practically a vertical strip covering 3 bins from 99 to 101 in the axis. This filter grid is applied to all the beams. This filter has the effect of removing most of the interfering signal at the expense of removing some energy of the signal. Note that the sidelobe leakage from the interfering source is also removed by this processing.

The (simulated) data are processed as follows. We select 10 blocks of overlapping data: each block covers 100 units of time and is spaced at 10 units of time apart from the next block. Each block of data is used to generate a 2D (time-frequency) beam intensity surface and its FFT spectrum. The contributions of the vertical striation components are removed from the FFT spectra, which are then integrated to generate the beam power. Figure 16(a) shows the conventional beam output at 10 time intervals (from Fig. 7). Figure 16(b) shows the beam power using the new processor. We note that between Figs. 16(b) and (a), the interference field

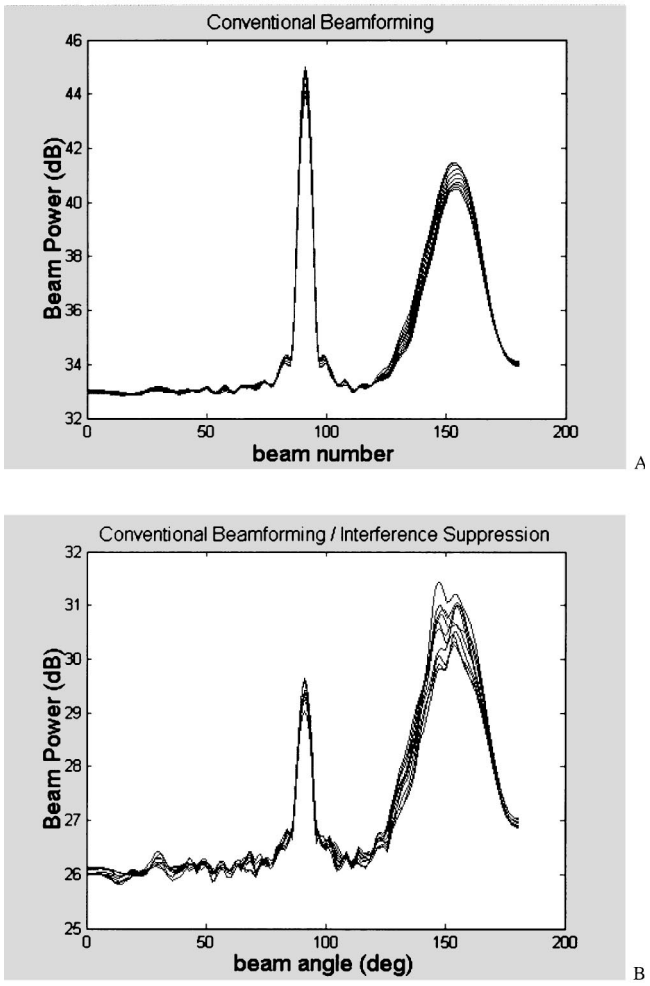


FIG. 16. The results of array processing which shows that the surface ship contribution has been suppressed in the array beam output (b) as compared with conventional beamforming output (a).

has been suppressed by 5–7 dB relative to the signal. Also we note that multipath arrivals are beginning to be resolved in Fig. 16(b).

Finally, we consider the case when the signal and interferer are located at the same bearing, i.e., 0°. This is the most difficult case for signal detection as the signal is usually masked by the interferer in the conventional beam outputs. But if the signal is close in and has a fast range rate, it will display a striation pattern which can be cued for signal presence. Figure 17 shows the broadside beam spectrogram as a function of time for the second scenario case (see above), where a loud surface ship is present in addition to the submerged signal. A nonvertical intensity striation pattern is suggested by Fig. 17. This pattern could suggest the presence of the submerged source. However, in practice, such a pattern could be distorted by the noise making it less noticeable. Figure 18 shows the FFT spectrum which could be used to discriminate against the noise and the distant, slow-moving (slow range-rate) surface ship. The presence of a nonvertical spectral component in the 2D FFT spectrum would be a cue for a close range, fast range-rate source.

V. SUMMARY AND DISCUSSIONS

This paper analyzes the beam intensity striation patterns as a function of frequency and range for a horizontal line

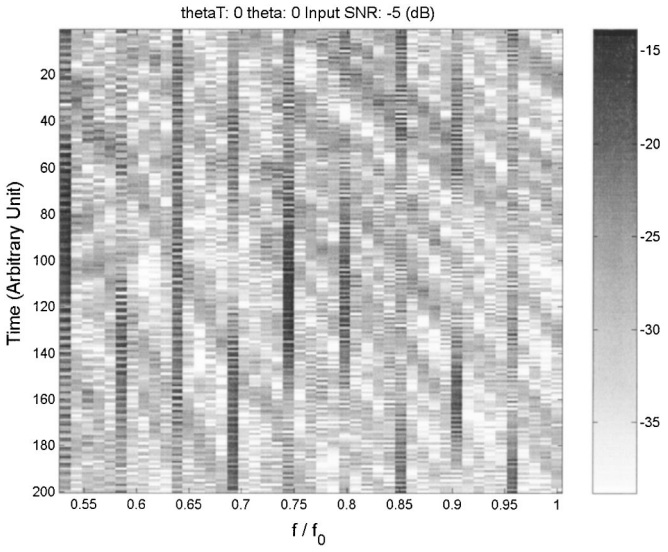


FIG. 17. Beam intensity as a function of frequency and time steered to the broadside direction, the bearing of both the target and the interference source. Same frequency and time as in Fig. 8 caption.

array obtained by conventional beamforming. The numerical examples clearly illustrate that the signal striation is better displayed by the beam intensity spectrum than the intensity at the element level due to the enhancement of signal by the array signal gain. Note that at the element level, the signal striation is often masked by an interfering field and noise when the signal is weak.

Analytical expressions were derived in this paper for the beam intensity as a function of frequency and range. Both analytical analysis and numerical simulations show that the beam intensities in the signal arrival directions exhibit the same striation pattern as that (predicted by the theory) for a single element. For a broadside signal, the beam intensity distribution is identical to that of a single element, enhanced by a constant signal gain. For a nonbroadside signal, the beam intensity level is modified by a frequency-bearing de-

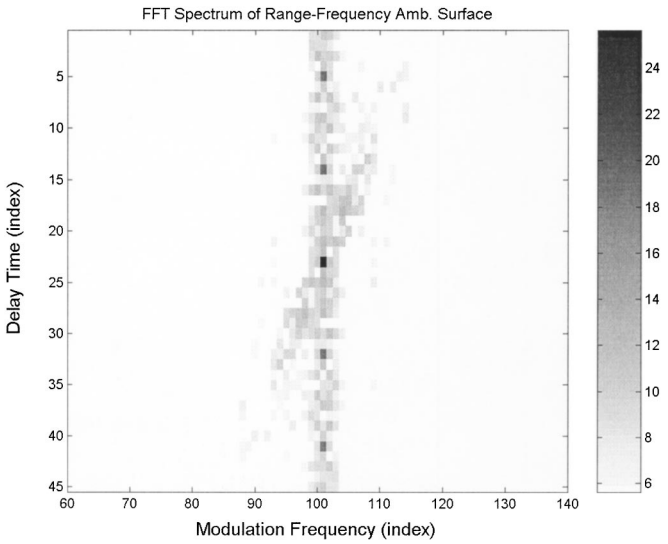


FIG. 18. Two-dimensional FFT spectrum of the beam intensity surface shown in Fig. 18. The *x* and *y* axes are the Fourier transform variables conjugate to time and frequency.

pendent signal gain. Despite the detailed intensity variations, the beam striation pattern (slope) remains the same as predicted by the theory. For signals originating from near endfire directions, the beam striation is better exhibited using the signal-arrival beam than the signal-look beam.

For nonsignal look beams (sidelobes), the beam intensity exhibits a frequency-modulated pattern associated with conventional beamforming (i.e., the modal beam sidelobe patterns) rather than the striation pattern of the signal.

Signal striation can be used as a cue for the presence of a fast range-rate, close range target among the noise and interference background, even when the signal and interference source are at the same bearing. Array processing can be improved by taking advantage of the difference in beam striations between the signal, interference source, and ocean noise. Numerical simulations in the SWellEX-96 environment showed a 5–7 dB suppression of the surface interferer. Actual improvement will depend on the array size, frequency, and features of the interference source. Although we have simulated the beam striations for a particular array configuration, we expect that the general features of our simulations and our conclusions will hold for other array configurations and design frequencies due to the invariant nature of the striation pattern. It should be pointed out that while the signal stayed in a fixed bearing for the above shown examples, the processing can be generalized to a source changing range and bearing at the same time. For the latter case, the beam spectrogram will be replaced by the scissorgram that follows the target.[15]

ACKNOWLEDGMENT

This work is supported by the Office of Naval Research.

[1] See, for example, I. Tolstoy and C. S. Clay, *Ocean Acoustics: Theory and Experiment in Underwater Sound*, 2nd ed. (American Institute of Physics, New York, 1987), Chap. 4.
[2] T. C. Yang, ''Dispersion and ranging of transient signals in the Arctic Ocean,'' J. Acoust. Soc. Am. **76**, 262–273 (1984).
[3] L. M. Brekhovskikh and Y. P. Lysanov, *Fundamentals of Ocean Acoustics*, 2nd ed. (Springer, New York, 1991), pp. 140–145, 148–152.
[4] S. D. Chuprov, ''Interference structure of a sound field in a layered ocean,'' *Acoustics of the Ocean: Current Status* (in Russian), edited by L. M. Brekhovskikh and I. B. Andreevoi (Nauka, Moscow, 1982), pp. 71–91.
[5] G. A. Grachev, ''Theory of acoustic field invariants in layered waveguide,'' Acoust. Phys. **39**, 33–35 (1993).
[6] 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).
[7] P. Gerstoft, G. L. D'Spain, W. A. Kuperman, and W. S. Hodgkiss, ''Calculating the waveguide invariant (beta) by ray-theoretic approaches,'' MPL Tech. Memo. TM-468, Marine Physical Laboratory, University of California, San Diego, December 2001.
[8] H. C. Song, W. A. Kuperman, W. S. Hodgkiss, T. Akal, and C. Ferla, ''A time-reversal mirror with variable range focusing,'' J. Acoust. Soc. Am. **103**, 3234–3240 (1998).
[9] A. M. Thode, W. A. Kuperman, G. L. D'Spain, and W. S. Hodgkiss, ''Localization using Bartlett matched-field processor sidelobes,'' J. Acoust. Soc. Am. **107**, 278–286 (2000).
[10] N. O. Booth, A. T. Abawi, P. W. Schey, and W. S. Hodgkiss, ''Detectability of low-level broad-band signals using adaptive matched-field processing with vertical aperture arrays,'' IEEE J. Ocean. Eng. **25**, 296–313 (2000).
[11] This assumption is only for the convenience of analytical calculations and is not needed in numerical simulations.
[12] N. O. Booth, P. W. Schey, and W. S. Hodgkiss, ''Broadband signal detection: Comparison of vertical aperture arrays using adaptive matched-field processing and a horizontal line array using adaptive plane-wave beamforming,'' J. Acoust. Soc. Am. **102**, 3170 (1997).
[13] M. B. Porter, ''The Kraken normal mode program,'' user's manual, SACLANT Undersea Research Centre, La Spezia, Italy, 1991.
[14] See, for example, T. C. Yang and K. Yoo, ''Modeling the environmental influence on noise vertical directionality in shallow water,'' J. Acoust. Soc. Am. **101**, 2541–2554 (1997).
[15] T. C. Yang, ''Motion compensation for adaptive horizontal line array processing,'' J. Acoust. Soc. Am. **113**, 245–260 (2003).

Acoustic scattering from a thermally driven buoyant plume revisited

John Oeschger
Coastal Systems Station, Naval Surface Warfare Center, Dahlgren Division, Panama City, Florida 32407-7001

Louis Goodman
School of Marine Science and Technology, University of Massachusetts Dartmouth, Intercampus Graduate School University of Massachusetts, 706 South Rodney French Boulevard, New Bedford, Massachusetts 02744

(Received 16 November 2001; revised 7 August 2002; accepted 4 December 2002)

Far-field weak scattering theory is applied to the case of high-frequency broad-bandwidth acoustic scattering from a thermally generated buoyant plume in a controlled laboratory environment. To first order, the dominant scattering mechanism is thermally driven sound-speed variations that are related to temperature deviations from ambient. As a result, the received complex acoustic scattering is a measure of the one-component three-dimensional Fourier transform of the temperature difference field measured at the Bragg wave number. The Bragg wave number vector is the difference between the scattered and incident wave vectors. Solving for its magnitude yields the Bragg scattering condition; this is the Fourier component of the plume variability that produces scattering. Results are presented for multistatic scattering from unstable and turbulent plumes using a parallel scattering geometry. The data justify application of the far-field weak scattering theory to the present case of a thermal plume. As a consequence, quantitative results on medium variability can be inferred using high-frequency broad-bandwidth acoustic scattering. Particular attention is given to the role of anisotropy of the variability of the scattering field in determining the validity of far-field Bragg scattering. [DOI: 10.1121/1.1542647]

PACS numbers: 43.30.Ft, 43.30.Vh [DLB]

I. INTRODUCTION

High-frequency broad-bandwidth acoustic scattering in water is of considerable interest to the physical and biological oceanographic communities. Since sound traveling through a medium will scatter due to impedance changes, it is of interest to explore whether it is possible to exploit these changes as a means to remotely measure medium variability such as fluctuations in the density, sound speed, or fluid velocity. Where these variations are such that weak scattering theory can be applied to describe the behavior of the sound, and that the scattering volume satisfies the far-field condition, quantitative spatial information of the medium variability can be obtained from the acoustic scattering process.[1,2]

Factors such as the presence of particulates and biologics in the water column also contribute to scattering and impede interpreting the received acoustic scattering signal just in terms of medium variability. Scattering from particulates can have similar and many times greater magnitude scattering strengths than that from impedance fluctuations of the medium. Thus, relying on signal level alone to determine the source of scattering can lead to interpreting incorrectly the acoustic data in terms of the type of scattering occurring. To avoid the uncertainty in knowing the source of the scattering, detailed knowledge of the scattering volume and scattering process is necessary. This information is also necessary to ground truth the acoustic scattering; however, such ground-truth information is often difficult to acquire.

Possibly the first calibrated *in situ* acoustic scattering experiment with accompanying environmental data was reported recently by Seim *et al.*[3] The experiment used downward-looking 120- and 200-kHz calibrated transducers in a salt-stratified tidal channel near Puget Sound, Washington. The environment contained energetic turbulent mixing events and was a near-ideal location to investigate acoustic scattering from turbulence and to make comparisons with environmental measurements. An advanced microstructure profiler was used to measure χ_T, the turbulent thermal variance dissipation rate and ε, the dissipation rate of turbulent kinetic energy. A Batchelor spectrum was employed to extend the measured temperature gradient spectrum beyond sensor capabilities. This allowed a spectral comparison between the environmental measurements and the acoustic scattering measurements at the same wave numbers. Seim *et al.*[3] found instances when the environmentally measured turbulence was large enough to account for the observed acoustic scatter. For the most part, however, the observed scattering strengths were higher than what the measured turbulent levels predicted. In most instances the scattering was likely due to particulates and biologics in the water column. Although great effort was taken to align the environmental and acoustic data sets (both spatially and temporally), unequal sampling volume sizes between the acoustically derived and oceanographically derived estimates of scattering strength led to the greatest source of uncertainty in these comparisons.

High-frequency acoustic scatter from biologics is capable of making biological biomass abundance estimates,

provided supporting information such as classification of scatterer, number density, and size distributions can be obtained to calibrate the acoustic data. More complex use of the acoustic scatter to infer scatterer dimensions and bulk material parameters is possible through application of various scattering models.[4–10] When applied in the forward direction, material parameters determined from independent measurements are used as inputs for the models and the scattering function thus obtained can be compared to that derived from the acoustic scattering. The amount of agreement depends upon the accuracy of the model, the accuracy of the material parameters used, and the ability to sample, both environmentally and acoustically, the same volumes. When the scatterer is known, and the scattering mechanism is well understood and accurately modeled, the acoustic data can be inverted to obtain quantitative information on the scatterers.[11–15] Similarly, high-frequency acoustic scatter from suspended sediments in the water column can be used to track regions of particulates, and in some instances, can yield quantitative information such as particle size and number density.[16–20]

Acoustic scatter from small-scale biologics and suspended sediments can be on the same order of magnitude as regions of intense turbulence.[3,4,21,22] A first step towards classification of the source of the acoustic scatter is to understand the scattering properties for each of the possible sources. This article focuses on examining the acoustic scatter from thermally generated turbulence in a controlled laboratory environment. The equation governing the acoustic scatter has been well established from first principles for a wave traveling through a region of medium variability.[23,24] The article reports on the application of far-field weak scattering theory to the case of high-frequency broadband acoustic multistatic scattering from a thermally generated buoyant plume.

Two previous investigations reported on the frequency and angular dependence of scattering for a cylindrically symmetric laminar plume[1] case and on the use of a multistatic scattering geometry for the unstable and turbulent plume cases.[2] The first set of measurements showed the acoustic scattering to be the result of thermally generated variability in the index of refraction. The laminar plume experiment resulted in a confirmation of the far-field weak scattering theory with the received complex acoustic scatter proportional to the one-component two-dimensional Fourier transform of the scattering field. An important consequence of far-field weak scattering theory is the Bragg scattering condition, which is described in detail in Sec. II. This condition can be used to expand the upper and lower limits of the spatial frequencies of the scattering field by employing simultaneous multistatic measurements with common Bragg wave numbers.

The Bragg scattering condition also provides an indirect means of verifying the approximations made to the scattering theory. For the specific case of the laminar plume, the far-field and weak scattering approximations used in the theory resulted in good agreement between the measured acoustic scatter and a model prediction that used independent temperature measurements.

For the more complex case of the turbulent plume, the environmental measurements required for intercomparison are difficult to obtain. This is because of the need to acquire simultaneous acoustic and environmental data over the same spatial region and at the resolution of order the wavelength of the acoustics. However, the validity of the constraint imposed by the Bragg scattering condition can be examined. The Bragg scattering condition leads to a technique whereby the received complex acoustic scatter obtained from different scattering angles can be compared in regions of equivalent/overlapping Bragg wave numbers. The far-field weak scattering theory predicts that the magnitudes of the acoustically estimated spectra obtained at different scattering angles are equal at equal Bragg wave numbers.

Unlike the case for the laminar plume, early results from the turbulent plume measurements did not confirm this prediction, thus suggesting a problem in the application of far-field weak scattering theory to the turbulent plume. An antiparallel scattering geometry experiment was designed to investigate specifically the cause of the failed prediction.[2] The results showed that wavefront curvature effects must be taken into account to describe properly the instantaneous scattering from the turbulent plume. Two undesirable consequences of this are: (1) that the received acoustic scatter no longer represents a one-component Fourier transform of the scattering field; and (2) that the Bragg scattering condition cannot be used as a substantiation of the scattering theory. Because the scattering from the turbulent plume is more similar to near-field than far-field scattering, the quantitative information of medium variability contained in the acoustic scatter appeared to be limited.

However, if instead of examining the instantaneous received pressure field associated with scattering from the turbulent plume, the ensemble average of the intensity is used, the Bragg scattering condition may be recovered. This results when the variability field has a sufficiently small degree of anisotropy. The scattered field can then be interpreted in terms of the wave number spectrum of the temperature field. It should be noted that for this case, since the scattering field is distributed throughout the scattering volume, the effective ''average'' spatial scale of scattering must be much smaller than the Fresnel radius.

In Sec. II the equation describing the acoustic scatter is given for the case of far-field weak scattering theory. This is followed by a description of the Bragg scattering condition. Subsections include discussions of near-field and beampattern effects on the scattering process and the usage of ensemble-averaged spectrum for isotropic variability to recover the Bragg scattering condition. Section III describes the experimental setup, laboratory procedures, and the measurements that are taken. In Sec. IV the results are given for two dynamically different thermal plumes. These are described as the unstable and turbulent thermal plumes. The summary and conclusions are given in Sec. V.

II. THEORETICAL BACKGROUND

High-frequency broad-bandwidth acoustics is well suited for studying volume variability in the ocean on mm to cm scales. The governing equation of motion is developed by considering the behavior of a sound wave as it travels through a homogeneous region and encounters a localized

anomaly.[1,2,23–26] This is expressed as the inhomogeneous Helmholtz equation with a source term representing the scattering due to the anomalous region. The source term is expressed as a function of three fundamental quantities: the relative compressibility, the density, and the fluid velocity. The case considered in this paper is that of acoustic scattering from a thermal plume where contributions to the acoustic scatter from density and fluid velocity variability can be neglected, this is shown in the Appendix. The compressibility or index of refraction changes are due to temperature variations of the thermally generated plume, to within an order of magnitude.

The complex acoustic scatter can be described analytically through the integral form of the Helmholtz equation. The integral contains the product of the pressure field, the relative compressibility, and the Green's function; the volume of the integral is defined by the source–receiver beam patterns. When the product of the relative compressibility, the acoustic wave number, and the length scale typical of the anomalous region within the scattering volume is small,[25] weak scattering theory is satisfied and the incident field, which can be measured at the volume, replaces the unknown total pressure field in the scattering volume. Simplification of the integral expression occurs when the scattering volume is in the far field of the source transducer. This means that in this circumstance the acoustic waves incident upon the scattering volume are approximated as planar.

As a consequence of using far-field weak scattering theory, the received complex acoustic scatter can be written in terms of a parameter "m" that denotes the spatial symmetry or dimensionality of the scattering process associated with the geometrical spreading of the scattered waves. This results by noticing the similarities that exist between the one-, two-, and three-dimensional Green's function solutions to the Helmholtz equation in the far field. The expression for the far-field approximation of the Green's function solution to the Helmholtz equation in m dimensions, where $m=1$, 2, or 3 can be written[27]

$$g_m(r,\mathbf{r}') = (\pi r)^{(1-m)/2} 2^{-(1+m)/2} k^{(m-3)/2} \times \exp[ikr - i\mathbf{k}_s\cdot\mathbf{r}' + i\pi(3-m)/4]. \quad (1)$$

The coordinate origin is defined at the center of the scattering volume; r is the distance from the origin to the receiver; $\mathbf{r}'$ is the vector from the origin to the scattering point; k is the acoustic wave number; and $\mathbf{k}_s$ is the scattered wave vector. Using the same theoretical development for the acoustic scatter as presented in a previous paper with the exception of replacing the three-dimensional Green's function [J. Acoust. Soc. Am. **100**, 1451–1462 (1996), Eq. (6)] with the m-dimensional form, the complex acoustic scatter in the frequency domain, neglecting the phase is

$$p_s(\omega) = p_o(\omega)(2\pi r)^{([1-m]/2)} k^{([1-m]/2)} \times \int B(\mathbf{x}')\mu(\mathbf{x}')\exp[-i\mathbf{K}\cdot\mathbf{x}']d^m\mathbf{x}'. \quad (2)$$

It is straightforward to verify that the resulting equations take on the correct forms when $m=1$, 2, or 3. $p_s(\omega)$ and $p_o(\omega)$ are the complex acoustic scatter measured at the receiver and the incident pressure field at the scattering volume; $\omega = ck$ is the angular acoustic frequency. $\mu = \Delta c/c$ is the relative index of refraction; B is the source–receiver beam pattern. K is defined as the Bragg wave number and is the magnitude of the difference between the scattered wave vector and the incident wave vector

$$K = |\mathbf{k}_s - \mathbf{k}_i|. \quad (3)$$

The magnitude of the Bragg wave vector is the product of twice the acoustic wave number and the sine of the scattering half-angle measured from the forward direction

$$K = 2k\sin(\theta/2). \quad (4)$$

Because of the importance of Eq. (4) in this paper, it is referred to as the Bragg scattering condition.

For a given source–receiver scattering geometry, the Bragg wave vector is aligned in a fixed direction. The notation is simplified by choosing a coordinate system such that the "x axis" is parallel to the Bragg wave vector. This geometry is illustrated in Fig. 1. Note the x axis is the bisector of the interior angle defined by the source to scattering volume and scattering volume to receiver vectors.

The Bragg wave vector number is the Fourier component of the scattering-producing field as indicated in Eq. (2). The focus of this report will deal with the integral in Eq. (2) and its interpretation. Consider

$$\phi_m(K) = \frac{1}{(2\pi)^m}\int B(\mathbf{x}')\mu(\mathbf{x}')\exp[-iKx']d^m\mathbf{x}', \quad (5)$$

where we will use either $m=2$ or 3 for the two- and three-dimensional scattering cases. The integral in Eq. (5) is written for the special case when the x axis is aligned with the Bragg wave number direction. Thus, for this case $\phi_m(\mathbf{k}) = \phi_m(k_x,k_y,k_z)$, with $k_x=K$, $k_y=k_z=0$. The beam-pattern

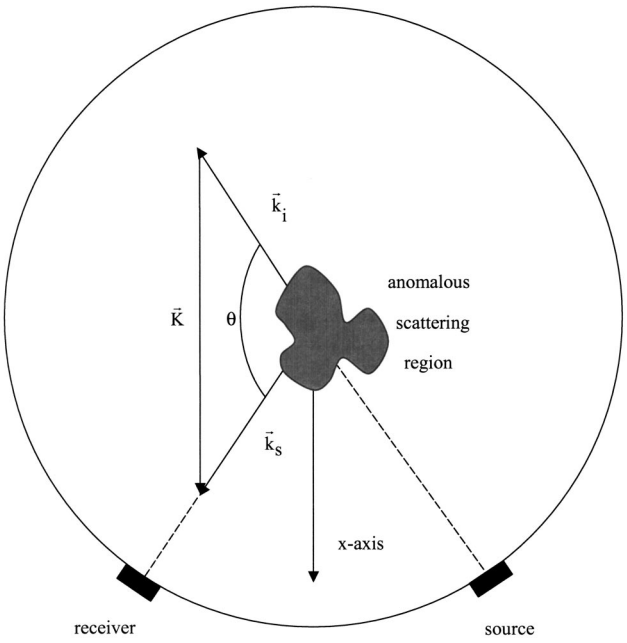


FIG. 1. Bragg wave number vector scattering geometry. Notice the x axis is defined as the bisector of the source to volume and volume to receiver cords. The scattering angle θ is measured from the forward direction.

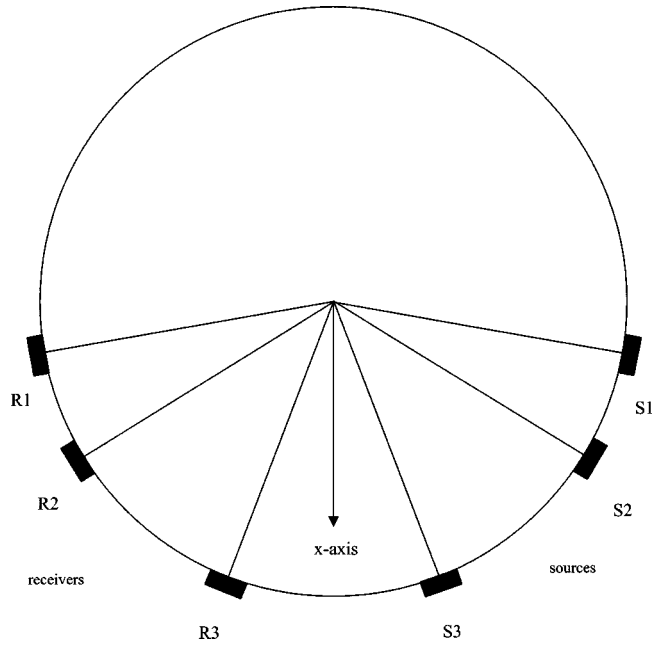


FIG. 2. Common Bragg direction multistatic scattering configuration. All source–receiver pairs have a common Bragg wave number vector direction and thus share the same bisecting x axis.

function B is taken as a smooth weighting to the integrand in Eq. (5), which is actually an approximation to the one-component, m-dimensional Fourier transform of the index of refraction fluctuations. The wave number resolution is proportional to the inverse of the effective scattering volume defined by B. The acoustical estimate of ϕ_m, denoted by γ_m, is determined by rearranging terms in Eqs. (2), (4), and (5) yielding

$$\gamma_m(K) = \frac{1}{(2\pi)^m} h(\omega)(2\pi r)^{(m-1/2)} \times \left(2 \sin\frac{\theta}{2}\right)^{([m+1]/2)} K^{-([m+1]/2)} = \phi_m, \qquad (6)$$

where $h(\omega) \equiv p_s(\omega)/p_o(\omega)$ is the transfer function of the scattering process. The relation between the Fourier transform of the temperature field and that of the index of refraction, the case considered in this paper, is $\phi = \alpha \phi_T$, where

$$\alpha = (1/c)(\Delta c/\Delta T).$$

Due to the finite bandwidth of available transducers, $p_o(\omega)$ can provide information over a limited range of acoustic frequencies. For a given acoustic bandwidth, the Bragg wave numbers determined from Eq. (4) and $k = \omega/c$ are similarly limited by the bandwidth of the incident acoustic pulse. This constraint limits the wave number information of Eq. (6). For example, the laboratory measurements reported in this paper use impulsive incident signals with 10-dB down points at 425 and 725 kHz. At backscattering, the available Bragg wave numbers for the stated frequency bandwidth range are between 3607 and 6156 rad/m. The significance of the Bragg scattering condition in Eq. (4) is seen by considering multiple bistatic measurements made in a common direction at a series of scattering angles. A schematic of this scattering geometry is given in Fig. 2 and is called the common Bragg direction configuration. In this arrangement the sources and receivers are placed on the perimeter of a circle and are directed into the axis of symmetry. Each source has a corresponding receiver. Note the parallel Bragg wave vector directions for each source/receiver pair.

One consequence of the broadband Bragg scattering condition is that multistatic scattering is capable of expanding the bandwidth of the acoustic estimate of the index of refraction (temperature) spectrum. Thus, when far-field weak scattering theory is satisfied, high-frequency broadband multistatic acoustic scattering from medium variability can be used to resolve more fully the wave number transform of the scattering field.

Providing a technique to ground truth these ideas is very difficult. At present there does not exist a quantitative technique to measure the three-dimensional structure of a temperature (and thus, for the case of a buoyant plume, the index of refraction) field over the domain of the scattering volume and at spatial resolutions of the order of or smaller than the Bragg scattering wavelength. The comparison that is performed in this paper is between the one-dimensional frequency spectrum obtained from a temperature probe and one derived from the acoustic scattering experiment. Since this is a limited testing of the theory, an alternative approach is to examine the constraints imposed by the consequence of Bragg scattering theory itself.

From the definition of ϕ_m given in Eq. (5) and estimated in Eq. (6) by γ_m, γ_m measured at different scattering angles corresponding to the same Bragg wave number must be identical, i.e.,

$$\text{If } K = K_i \cap K_j, \text{ then } \phi_i(K) = \phi_j(K). \qquad (7)$$

That is, $\phi_i(K) = \phi_j(K)$ for all K which are common to both K_i and K_j; here, the subscripts denote measurements made at the ith and jth scattering angles. Equation (7) forms a basis through which the far-field weak scattering theory can be at least indirectly verified when applied to the case of the buoyant plume.

A. Near-field effects

An earlier paper showed that predictions made by the far-field weak scattering theory do not always hold true when applied to the laboratory case of scattering from a thermally generated buoyant plume.[2] If $K = K_i \cap K_j$, and $\phi_i(K) \neq \phi_j(K)$, then there must be a breakdown in one of the two assumptions used to describe the acoustic scattering. These two assumptions are: (1) the far-field approximation; and (2) weak scattering theory. The condition for weak scattering theory to be valid is given by $\mu kD \ll 1$, where μ is a characteristic value of the index of refraction on the scale of the scattering, k the acoustic wave number, and D the characteristic value of the length scale of the scattering.[25] For the turbulent plume considered here, values of these parameters are $\mu = 10^{-3}$, $k = 3*10^3$ rad/m, $D = 1*10^{-2}$ m, such that $\mu kD = 0.03$. The condition for weak scattering theory to be valid is satisfied. This suggests the far-field approximation is the likely source for the inequality in $\phi_i(K) \neq \phi_j(K)$.

When the far-field approximation is not satisfied, higher-order terms need to be included in the exponential of Eq. (2).

TABLE I. Table of frequency-dependent values of terms relevant to this section. Acoustic frequency f, acoustic wavelength $\lambda = c/f$, sound speed $c = 1500$ m/s, acoustic wave number $k = 2\pi/\lambda$, radius of first Fresnel zone $r_F = \sqrt{r\lambda/2}$, $r = 25$ cm, radius of beam at scattering volume (at -10 dB) $r_b = 5.4r/(ka)$, a = radius of transducer = 1.9 cm (0.75 in.), number of Fresnel zones $n_F = 2r_F/(r\lambda)|_{r_F = r_b}$, first axial near-field null coming in from infinity $r_{\min} = a^2/(2\lambda) - \lambda/2$, first axial maximum coming in from infinity $r_{\max} = a^2/\lambda - \lambda/4$; and also, in this paper, $r/a = 13.1$.

f kHz	λ mm	k rad/m	r_F cm	r_b cm	n_F	$(r_F/r_b)^2$	ka	$r_{\min}$ cm	$r_{\max}$ cm
425	3.5	1780	2.1	4.0	3.6	0.3	34	5.0	10.2
525	2.9	2199	1.9	3.2	2.9	0.3	42	6.2	12.6
625	2.4	2618	1.7	2.7	2.4	0.4	50	7.4	15.0
725	2.1	3037	1.6	2.3	2.1	0.5	58	8.7	17.5

Consider, for simplicity, the case of three-dimensional scattering. Results for the two-dimensional scattering case are entirely analogous. Equation (2) then becomes

$$p_s(\omega) = p_o(\omega)(2\pi)^2 r^{-1} k^2 \Gamma, \tag{8}$$

where

$$\Gamma(K) = \frac{1}{(2\pi)^3} \int B(\mathbf{r}')\mu(\mathbf{r}')\exp[-iKx' + ik\zeta]d\mathbf{r}'$$

$$= \frac{1}{(2\pi)^3} \int_V \mu(\mathbf{r}')\exp[-iKx' + ik\zeta]d\mathbf{r}', \tag{9}$$

and the phase correction term, ζ, to lowest order is given by $\zeta = (x'^2\cos^2(\theta/2) + y'^2\sin^2(\theta/2) + z'^2)/r$. Note that γ_m with $m = 3$ of Eq. (6) is the acoustic estimate of Γ if $\zeta = 0$. The notation of subscripting the second integral in Eq. (9) with V indicates that the beam-pattern weight B is used over the volume domain of integration. The beam pattern acts like a spatial Hanning window in this integral. Note that $\Gamma(K)$ no longer represents a Fourier transform of the index of refraction field, because of the presence of the term ζ, and thus that the multistatic scattering configuration in a common Bragg wave number direction can no longer be employed to extend the effective bandwidth of the estimated Fourier transform of the scattering field. The quadratic terms of ζ represent wavefront curvature or near-field-like correction terms to the far-field approximation.

B. On the nature of Bragg scattering from a turbulent field

Previous unpublished results for multistatic scattering in a common scattering direction from a turbulent plume showed a breakdown of the prediction of the Bragg scattering condition given in Eq. (7), i.e., $\phi_i(K) \neq \phi_j(K)$ when $K = K_i \cap K_j$. Typically, the far field of a baffled piston is defined as $r/a \gg 1$ and $r/a \gg ka$, where r is the range to the field point, and a is the radius of the acoustic source.[28] Approximating the acoustic beam radius at the volume as $r_b = r/(ka)$ and the Fresnel radius[29] as $r_F = \sqrt{r\lambda/2}$, then the previous inequalities can be restated as $r_b \gg \lambda$ and $(r_F/r_b)^2 \ll 1$. This last inequality implies that when the far-field condition is satisfied, there are many Fresnel zones within the volume defined by the acoustic beam. Table I shows values of the various parameters for the experiment described in this paper, which uses a cylindrical baffled piston.

The presence of many Fresnel zones on the incident wave front may or may not be desired, depending upon the relation between the Fresnel radius and the spatial extent of the scattering mechanism in the direction of the Fresnel radius (i.e., perpendicular to the direction of the vector Bragg wave number). For instance, consider measuring the scattering characteristic of a localized, spherical point-like scatterer. In this case, as the range to the scattering volume increases, the Fresnel radius increases, and when compared to the constant length scale of the scatterer, the wavefront over the scattering region appears more planar than spherical. Contrast this type of point-like scatterer to one where the scattering mechanism is distributed throughout the entire scattering volume, such as a turbulent plume with index of refraction variability. In this case, as the scattering volume is located further into the far field, the wavefront incident upon the entire scattering volume appears more spherical than planar, and in our case more spherical to individual (thermal) turbulent eddies which produce the scattering. This is a result of the beamwidth at the volume increasing with range faster than the Fresnel radius increases with range. This implies that scattering from a distributed field of scatterers (such as from the thermal plume used in the measurements in this article) may be in the near field. This might invalidate the usage of the far-field approximation, the Bragg scattering condition, and the ability to obtain an estimate of the Fourier transform of the scattering field. Far-field scattering requires that the size of the scatterer in the direction of the Fresnel radius be smaller than the Fresnel radius. Thus, the degree of symmetry of a scatterer or, in the case of statistical quantities such as turbulent spectra, the degree of anisotropy of an ensemble of scatters will play an important role in determining whether the Bragg scattering condition can be used to develop a Fourier integral relationship between the received acoustic field and the scattering field. These issues will be discussed in detail in the next section. Note aspects of this have also been discussed by one of the authors in a previous paper.[23] If l is a characteristic scale of the scattering in the direction perpendicular to the Bragg wave number, then the criterion for far-field Bragg scattering is $l \ll r_F$.[30] The correlation length of the temperature anomalies for the turbulent plume considered in this paper in the direction perpendicular to the plane of scattering (the "z" direction) and thus along a Fresnel radius can be qualitatively estimated to range between $0.1\text{ cm} < l < 2.0\text{ cm}$. Thus, for the experiment to be described in this paper, l can be of order of the Fresnel radius

(see Table I) and eddies of the turbulent thermal plume may not satisfy the criterion of having a sufficiently small spatial scale in the direction of a Fresnel radius at any instant of time. This may be why the previous (unpublished) data set indicated a breakdown of the Bragg scattering condition. However, if the variability of the temperature field eddies within the thermal plume are turbulent-like at the scale of the Bragg wavelength, then it is expected that the ensemble-averaged received pressure field may result in the far-field Bragg scattering condition being satisfied and a Fourier integral recovered. This can occur because in a turbulent field, although individual turbulent eddies may have a particular orientation, on average, their orientation is more random and averaging is expected to bring down the degree of asymmetry or anisotropy. Note that in classical turbulence theory at scales of the inertial subrange and smaller[31] the three-dimensional spectrum of the field is isotropic even though individual eddies are not necessarily (and indeed in general are not) spherically symmetric. See, for example, numerical studies on turbulence in Refs. 32–34. Thus, for isotropic turbulence the far-field scattering condition and the Bragg scattering condition can then be recovered and the spectrum of the temperature field estimated acoustically. Realistic turbulent-like fields may have anisotropy, but it is still expected that the degree of anisotropy will be less than the degree of asymmetry of an individual instantaneous eddy. In the next section criteria for the degree of maximum anisotropy for Bragg scattering to be valid are derived. The reader again is also referred to Ref. 23 for a discussion of this issue. Thus, for a turbulent field the scale or wave number of the scattering will play a critical role in determining this issue of isotropy. If the Bragg wave number is of the order and/or greater than a characteristic inertial subrange wave number[31] it is expected that isotropy or near-isotropy should hold. Alternatively, as we will show in the next section, acoustic scattering provides a mean of assessing whether in fact the turbulent field has this level of isotropy.

C. Recovering the Bragg scattering condition

Using Eq. (9), consider the ensemble average of received acoustic intensity

$$\begin{aligned}\Psi &= \langle \Gamma' \Gamma'^{*} \rangle \\ &= \frac{1}{(2\pi)^6} \int_V \int_V d\mathbf{r}'' d\mathbf{r}' \langle \mu'(\mathbf{r}'') \mu'(\mathbf{r}') \rangle \\ &\quad \times \exp[-iK(x''-x') + ik(\zeta''-\zeta')] \\ &= \frac{1}{(2\pi)^6} \int_V \int_V d\mathbf{r}\, d\tilde{\mathbf{r}}\, R(\mathbf{r}) \exp\left[-iKx \right. \\ &\quad \left. + \frac{2ikx\tilde{x}\cos^2\theta/2}{r} + \frac{2iky\tilde{y}\sin^2\theta/2}{r} + \frac{2ikz\tilde{z}}{r}\right], \end{aligned} \tag{10}$$

where the variables in the second equation above have been changed according to $\mathbf{x}=\mathbf{x}''-\mathbf{x}'$ and $\tilde{\mathbf{x}}=(\mathbf{x}''+\mathbf{x}')/2$, and $\langle \cdots \rangle$ indicates an ensemble average. In Eq. (10), note that we have defined the fluctuating pressure field in terms of the fluctuating index of refraction field by use of the prime notation, namely $\Gamma'(\mathbf{r}',\tau)=\Gamma(\mathbf{r}',\tau)-\langle \Gamma(\mathbf{r}',\tau)\rangle$, and $\mu'(\mathbf{r}',t)=\mu(\mathbf{r}',\tau)-\langle \mu(\mathbf{r}',\tau)\rangle$. Using the definition of the index of refraction spectrum as $R(\mathbf{r}''-\mathbf{r}')=\int \Phi(\boldsymbol{\kappa}) \exp[i\boldsymbol{\kappa}\cdot(\mathbf{r}''-\mathbf{r}')]d\boldsymbol{\kappa}$, Eq. (10) leads to

$$\Psi = \frac{1}{(2\pi)^3} \int_V d\tilde{\mathbf{r}}\, \Phi\left(K + \frac{2k\tilde{x}\cos^2\theta/2}{r}, \frac{2k\tilde{y}\sin^2\theta/2}{r}, \frac{2k\tilde{z}}{r}\right). \tag{11}$$

Note that Eq. (11) can be interpreted as a spatial average of the wave number spectrum centered at the Bragg wave number. At this point no assumption on the nature of the spatial structure of Φ has been made. Let us write the integrand of (11) as

$$\Phi(K+\Delta k_x, \Delta k_y, \Delta k_z), \tag{12}$$

where

$$\Delta k_x = \frac{2k\tilde{x}\cos^2\theta/2}{r}, \tag{13a}$$

$$\Delta k_y = \frac{2k\tilde{y}\sin^2\theta/2}{r}, \tag{13b}$$

$$\Delta k_z = \frac{2k\tilde{z}}{r}. \tag{13c}$$

Equation (12) can be interpreted as the wave number spread on the spectrum Φ due to the effect of the Fresnel phase terms. If these terms can be neglected, i.e., $\Delta k_x \approx \Delta k_y \approx \Delta k_z \approx 0$, then use of Eq. (12) in (11) yields

$$\Psi = (2\pi)^3 V \Phi(K,0,0) = (2\pi)^3 V \Phi(K). \tag{14}$$

Thus, the Bragg scattering condition is recovered and the received spectrum is proportional to the spectrum of media variability producing the scattering. Clearly, Δk_x, Δk_y, Δk_z in Eq. (12) represents the wave number shift (and on averaging the wave number smearing) of the integral of Eq. (11). The criterion for neglecting these quantities is easily established since the typical maximum spatial extent of the scattering volume in any one direction is approximately $2r\theta_b$, where $\theta_b \cong 2.7/(ka)$ is a characteristic half-beamwidth angle; then, the criterion for neglecting the first term, Eq. (13a), is $\Delta k_x \ll K$

$$2\theta_b \cot(\theta/2)\cos(\theta/2) \ll 1. \tag{15}$$

Except for cases of very near-forward scattering, when $\theta_b \approx \theta/2$, Eq. (15) is well satisfied for sufficiently small values of θ_b. This will be the case for the experiment described below. Criteria for establishing the neglect of Δk_y, Δk_z are dependent upon the spatial asymmetry of the variability along the three spatial axes.

Let α_y and α_z be the degree of anisotropy of Φ in the "y" and "z" direction relative to the "x" axis, and let us assume that the x axis of the Bragg direction is aligned in the direction of the smallest principal axis of symmetry; then, both $\alpha_y \leqslant 1$ and $\alpha_z \leqslant 1$. It follows that $\Phi(K,k_y,k_z) \approx \Phi(K,0,0)$ if

$$\Delta k_y \ll \alpha_y K, \tag{16a}$$

and

$$\Delta k_z \ll \alpha_z K, \tag{16b}$$

and Eq. (14) follows, which recovers the Bragg scattering condition for the wave number spectrum. To establish the criterion for this, substitution of Eqs. (13a) and (13b) into (16a) and (16b) yields

$$2\theta_b \sin(\theta/2) \ll \alpha_y, \tag{17a}$$

$$2\theta_b \ll \alpha_z \sin(\theta/2). \tag{17b}$$

For the scattering angles used in the experiment of the paper, these conditions are approximated by

$$\theta_b \ll \alpha_{y,z}. \tag{18}$$

Thus, if the degree of anisotropy parameter $\alpha_{y,z}$ is not smaller than θ_b the Bragg scattering condition can be applied. This result also holds for the two-dimensional scattering case associated with (two-dimensional) turbulent-like variability.

To obtain the acoustical estimate of the three-dimensional case we substitute γ_3 from Eq. (6) into Eq. (14) using the definition of the wave number spectrum to obtain for the acoustically derived three-dimensional wave number the estimate of index of refraction, namely

$$\Psi_{3D}(K) = \frac{1}{V(2\pi)^3} |h(\omega)|^2 (2\pi r)^2 (2\sin(\theta/2))^4 K^{-4}, \tag{19}$$

where V is the scattering volume defined by the beam-pattern intersection. Analogously, we can obtain the estimate for the two-dimensional scattering case, which can be shown to be

$$\Psi_{2D}(K) = \frac{1}{A(2\pi)^2} |h(\omega)|^2 (2\pi r)(2\sin(\theta/2))^3 K^{-3}, \tag{20}$$

where A is the area of the two-dimensional scattering field. For the case of the baffled pistons used in the experiments defined below, we take $V = L\ A$ with $L = 5.4\ r/(ka)$, the vertical extent of the scattering volume defined to the 10-dB down point, where r is the range from the receiver to the scattering volume and a is the transducer radius. The ratio of the three-dimensional estimate to the two-dimensional estimate is then found to be

$$\frac{\Psi_{3D}(K)}{\Psi_{2D}(K)} = 0.2a. \tag{21}$$

Thus, the three-dimensional estimate will scale with Bragg wave number the same as the two-dimensional estimate and differ by a factor of order a. Equation (21) will have a very important consequence in comparing the acoustically estimated index of refraction spectrum of the plume with that derived from a direct measurement of index of refraction (temperature). For Eqs. (19), (20), and (21) to be valid, the condition in Eq. (18) must hold.

III. EXPERIMENTAL SETUP

The laboratory consists of a $2 \times 2 \times 1.2$-m polypropylene tank filled with fresh tap water and filtered down to 5 microns. Glass windows are located on each side of the tank, allowing use of a laser shadowgraph system to image the thermal plume and qualitatively monitor its condition. The heating element used to generate the plume is inside a cylindrical stainless-steel housing having outside diameter of 1 cm, 3.8 cm high, elevated 6 cm above a bottom mounting plate and located 25 cm below the horizontal scattering plane. The input power to the heating element is controlled with a Variac with a maximum output power of 50 watts. The acoustic volume is centered on the plume axis by using a ring assembly system to mount the transducers. The assembly consists of a plate having a 50-cm-diameter inner radius around which transducers are placed in 10-deg increments. For this work, three pairs of reciprocal source–receiver transducers were used, each having 10-dB down points at 425 and 725 kHz. The transducers have a crystal diameter of 1.5 in. and are well modeled by baffled pistons.

The transmitted acoustic pulses are generated using an arbitrary waveform generator consisting of a single-cycle pulse with center frequency at 500 kHz and amplified with 2-kW amplifiers. Each of the three source transducers is stimulated separately with a 1-ms time lag between transmissions, thus sufficiently reducing tank reverberation yet still allowing interrogation of essentially the same scattering field. Since the nominal plume velocity in the vertical is 3 cm/s, typical vertical displacement over 2.0 ms (the time delay from the stimulus of transmitter 1 until transmitter 3 is fired) is approximately 0.06 mm, which has a negligible effect on the results. Each receiving channel is range gated such that for each ping the acoustic scatter from the plume is centered in the 512 samples that are digitized at 5 MHz. The system repetition rate is 67 Hz, and 2048 waveforms per channel are collected. The high repetition rate allows for estimating one component of the turbulent velocity using a coherent acoustic Doppler technique and is a subject of continuing research. Incident acoustic signals at the scattering volume are measured by placing the receiving transducer directly facing the source transducer across the ring.

A Seabird SBE 3-01/F fine-structure temperature probe is also placed on the plume axis 7.5 cm above the horizontal acoustic scattering plane. The probe has a measured e-folding time of 67 ms. Temperature time series measurements are sampled at 60 Hz and made simultaneously with the acoustic scattering measurements. Because of the inhomogeneous nature of the plume turbulence, the spectral estimates made 7 cm above the scattering volume differed from the estimate made at the center of the volume. Thus, temperature measurements are also made both before and immediately after the acoustic measurements, with the temperature probe, however, moved to the center of the scattering volume.

IV. RESULTS

The data collected for the measurements made in the common Bragg scattering configuration (Fig. 2) direction have scattering angles of 80, 120, and 160 deg measured, respectively, from the forward direction for each of the three source–receiver pairs used. The fluctuations about the mean are used for the analysis and are determined by removing the mean waveform for each channel from each waveform of the

respective channel. That is, let the nth ping received on the ith channel be represented by $p_i(t_j,\tau_n)$; the variable t_j represents the time of the sampled points per ping, while the τ_n is the repetition or ping interval. For this paper, $i=1$, 2, or 3; $j=1,...,512$; and $n=1,...,2048$. The t_j can be converted to a spatial coordinate through the relation $x_{i,j}=ct_j/(2\sin\theta_i/2)$, while the Bragg wave numbers for each channel are determined by $K_i=2k\sin(\theta_i/2)$. The mean waveform for the ith channel is then expressed by $\overline{p_i}(t_j)=\langle p_i(t_j,\tau_n)\rangle$, so that the fluctuations of the ith channel are represented by $p_i'(t_j,\tau_n)=p_i(t_j,\tau_n)-\overline{p_i}(t_j)$. The angle brackets represent an ensemble average over the $N=2048$ waveforms taken for each experiment. It is assumed that μ does not vary over time scales t_n which are of order tens of microseconds. Note that the Fourier transform of p' on t is Γ'.

A. Unstable plume

For the purposes of this paper, the buoyant thermal plume has three different dynamical regimes denoted by laminar, unstable, and turbulent flows. These regimes are obtained by varying the input power to the heating element. Nominally, the plume remains laminar at the height of the scattering plane for an input power level less than 25 W. An unstable plume exists for input power levels between 25–35 W and turbulent for power levels greater than 35 W. It should be noted, however, that these power levels might shift by as much as 50% in either direction over the period of hours. For the data presented in this section, the unstable plume is generated with an input power level of 35 W. The unstable plume for this case is characterized by a slow oscillation in

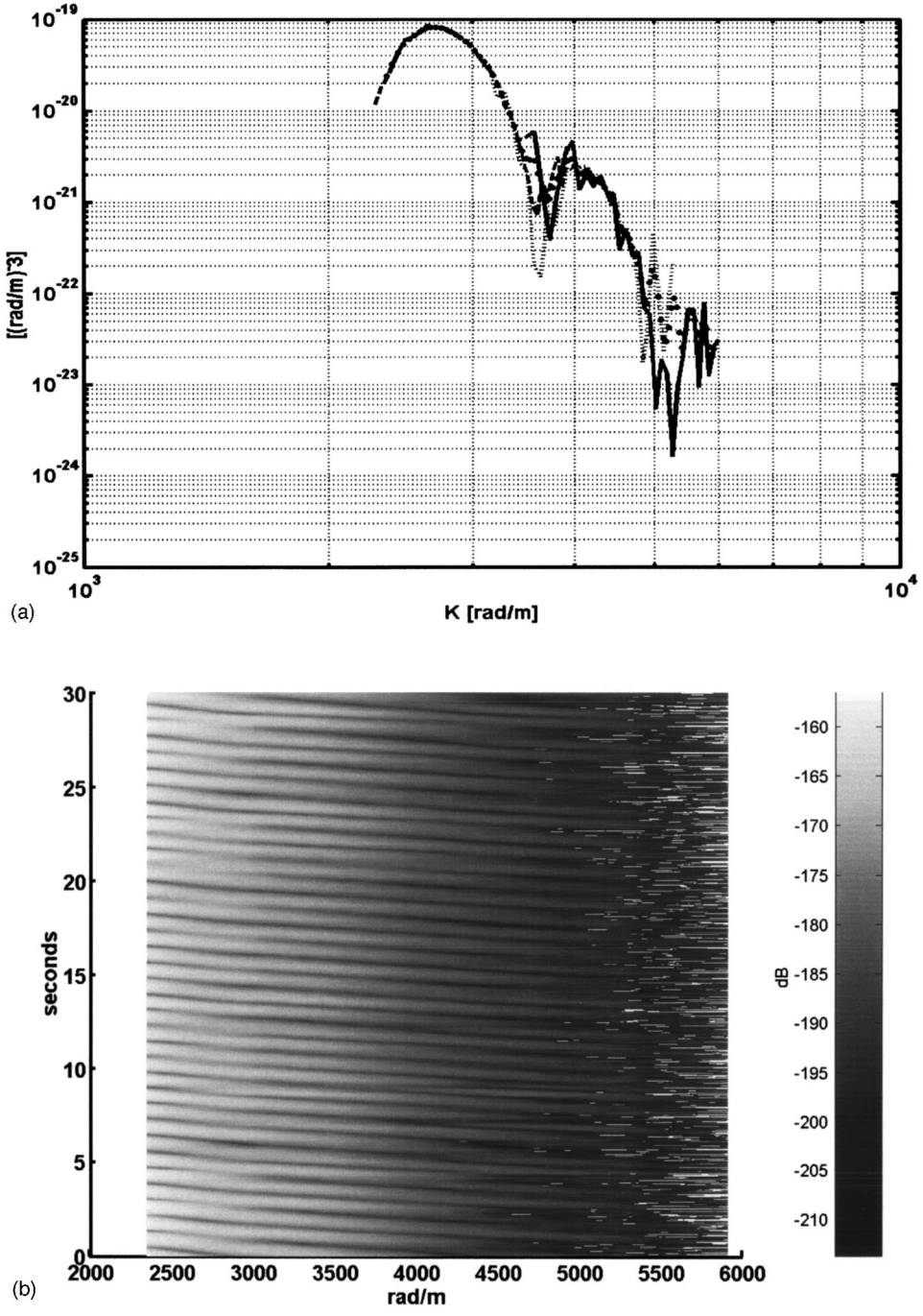


FIG. 3. (a) Common Bragg wave number spectrum of index of refraction fluctuations for the unstable plume at $t=22.575$ s. The spectrum is formed by combining data from three separate receiver channels corresponding to scattering angles of 80 deg (dashed), 120 deg (dotted), 160 deg (solid). The black dots are an interpolation of the three curves over the entire range. (b) Time series of the interpolated common Bragg wave number spectrum of index of refraction fluctuations [black dots of (a)] for the unstable plume case. (c) Temporal mean wave number spectrum calculated from (b), the case of the unstable plume.

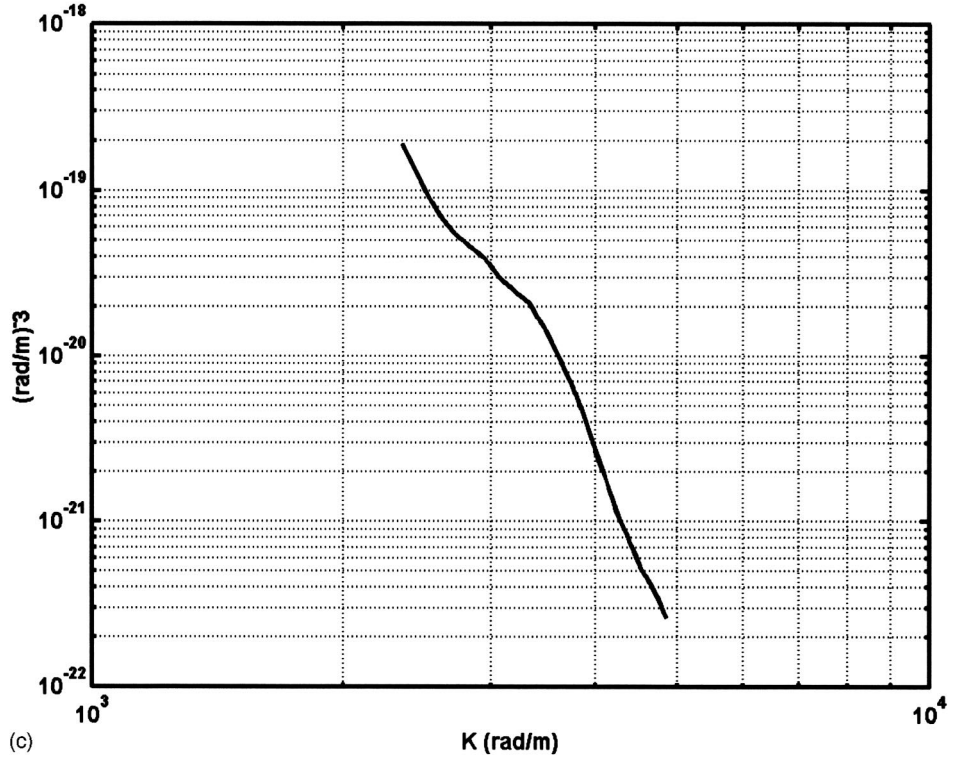


FIG. 3. (*Continued.*)

the horizontal plane, giving rise to a sinusoidal-like plume structure in the vertical direction with a nominal vertical wavelength of 5 cm.

Consider the three-dimensional spectrum of the index of refraction field, which we will estimate with Ψ_{3D}, using Eq. (19) in the common Bragg scattering configuration. Figure 3(a) shows Ψ_{3D} for an unstable plume at 22.575 seconds into a 30.72-s measurement run. The dashed, dotted, and solid lines are the contributions to Ψ_{3D} from the 80-, 120-, and 160-deg scattering angle data sets, respectively. An estimate of Ψ_{2D}, using Eq. (20), will yield the exact same figure but differing by the factor $0.2a$ as discussed and shown by Eq. (21). [Note that the units would be different also, i.e., $(\text{rad/m})^{-2}$ instead of $(\text{rad/m})^{-3}$.] The strong degree of overlap suggests that the Bragg scattering condition is well satisfied, but at this stage of analysis we cannot say if scattering from the unstable plume is a process which is more three-dimensional than two-dimensional. We will return to this point later.

Returning to the very good agreement in the spectral magnitude in the regions of overlapping Bragg wave numbers, these occur between channels 1 and 2 for wave numbers from 2000 to 3000 rad/m and for channels 2 and 3 from 3500 to 5500 rad/m. Noteworthy is the presence of deep nulls in the wave number spectrum, indicative of sharp temperature gradients in the thermal plume.

A contour plot of $\Psi_{3D}(K,\tau)$ versus Bragg wave number and time is shown in Fig. 3(b). All three channels of data are combined into one matrix, sorted by wave number component, and interpolated over an evenly spaced set of wave numbers corresponding to the range of Bragg wave numbers available from the scattering measurement. Note the smooth variation in the magnitude of the spectrum as a function of wave number. This is seen throughout the 30-s time series measurement and indicative of the success of the common Bragg wave number comparisons. The periodicity in the spectrum is evident as expected for an unstable plume. It is interesting to examine the slope of constant contours in the spectrum, indicative of constant null spacing in the spectrum throughout the time series. There results a typical value of $\Delta K_x/\Delta\tau = 1250$ rad/(ms) to 2500 rad/(ms). Laser shadowgraph observations suggest in the vertical direction ($\lambda_z = 5$ cm, $T = 1.7$ s) a value of $\Delta K_z/T = 74$ rad/(ms). The ratio of these two quantities results in an estimate for the length scale of horizontal variability between $l_x = l_z/17 = 2.9$ mm and $l_x = l_z/34 = 1.5$ mm.

If the temperature profile is modeled as a cylinder, then the solution for the scattered pressure will vary as $J_1(Kb)$, where J_1 is the Bessel function of order one, and b is the radius of the cylindrical plume profile. Except for the first pair, the spacing between the first few zeros of $J_1(x)$ is constant ($\Delta x_i = 3.83, 3.19, 3.15, 3.15, 3.15$). Given an estimate for b of 2.5 mm from above, the first Δx can be safely ignored since that null occurs at $K = 1530$ rad/m, which is outside of the available Bragg wave number bandwidth. From Figs. 3(a) and (b), an estimate of the plume width is found by measuring the null spacing in wave number and calculating $b = \Delta x_3/\Delta K$. Typical ΔK values between nulls range from 1200 to 2000 rad/m; these correspond to estimates of b between 1.6 and 2.6 mm. These estimates are in close agreement with the value of 2.4 mm associated with the laminar plume described in an earlier paper.[1]

Figure 3(c) is the mean three-dimensional wave number spectrum of the index of refraction fluctuations for the unstable plume. The spectrum includes data from all three channels and indicates the good agreement between the re-

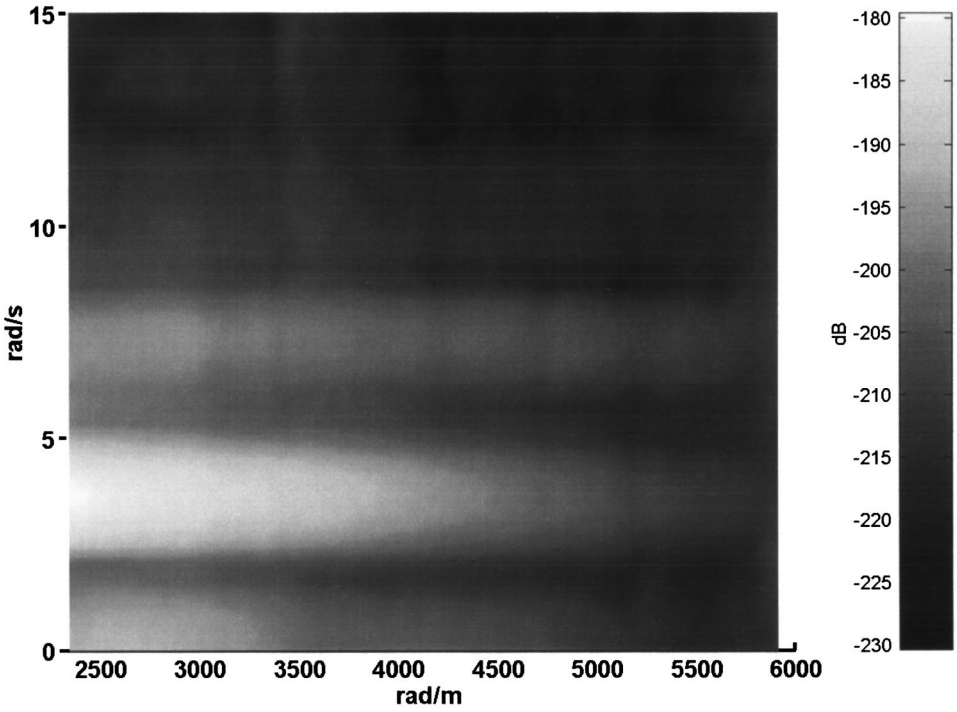


FIG. 4. Wave number/frequency spectrum of index of refraction fluctuations for the unstable plume case. Same data as those used in Fig. 3(b).

gions of overlapping Bragg wave numbers. For the unstable plume, the spectrum in the given wave number range falls off rapidly at a power law of K^{-9}.

The wave number frequency spectrum, defined by

$$\Psi_{3D}(K,\omega)=\frac{(2\pi)^4}{TV}\langle\gamma_3'(K,\omega)\,\gamma_3'^{*}(K,\omega)\rangle, \tag{22}$$

where $\gamma_3'(K,\omega)$, is the Fourier transform in time of $\gamma_3'(K,\tau)$ from Eq. (6), K is the Bragg wave number, V the scattering volume, and T the total time over which the data were sampled. Figure 4 clearly shows the periodicity of the unstable plume over the entire bandwidth of wave numbers with the peak magnitude occurring between 3 and 4 rad/s and a second local maximum around 7 rad/s. From the relationship of $\omega=\nu k_z$ and using 3.5 rad/s for ω and 3 cm/s for ν, the value for the wavelength of the oscillation in the vertical direction is 5.4 cm. This is in close agreement with laser shadowgraph observations.

To compare the acoustically derived results with those from the temperature probe we need to convert the acoustic wave number frequency spectrum to a frequency spectrum, since the temperature probe is limited to a point time series measurement. An acoustically estimated one-dimensional frequency spectrum $\Psi(\omega)$ is formally related to the three-dimensional wave number/frequency spectrum by

$$\Psi_{3D}(\omega)=\int d^3\mathbf{k}'\Psi_{3D}(\mathbf{k}',\omega), \tag{23}$$

and to a two-dimensional wave number/frequency spectrum by

$$\Psi_{2D}(\omega)=\int d^2\mathbf{k}'\Psi_{2D}(\mathbf{k}',\omega). \tag{24}$$

Since we only have wave number information in one direction we need to make an assumption about the statistical nature of the variability in the other two spatial directions. The assumption that we make (under the criteria developed in Sec. II) is that of spatial near-isotropy over the Bragg wave number range (which corresponds to mm range spatial scales), which then results in

$$\Psi_{3D}(\omega)=(4\pi)\int_{k_1}^{k_u}dk'\,k'^2\Psi_{3D}(k',\omega), \tag{25}$$

where k_l and k_u are upper and lower limits of integration defined by the finite bandwidth used in the experiment. If a two-dimensional scattering process were assumed and Eq. (24) utilized again with the same assumption, then the frequency spectrum would be

$$\Psi_{2D}(\omega)=2\pi\int_{k_1}^{k_u}dk'\,k'\Psi_{2D}(k',\omega). \tag{26}$$

Using Eq. (21), such that

$$\Psi_{3D}(K)=0.2a\Psi_{2D}(K), \tag{27}$$

results in

$$\Psi_{3D}(\omega)\approx K^{*}a\Psi_{2D}(\omega)\gg\Psi_{2D}(\omega), \tag{28}$$

where K^{*} is a characteristic wave number whose value is expected to be between k_l and k_u, and depends on the details of the integration. Thus, if one uses a three-dimensional estimate, the estimate $\Psi_{3D}(\omega)$ will be much larger than that from a two-dimensional estimate $\Psi_{2D}(\omega)$ by the factor $K^{*}a$. This can be understood by the fact that a two-dimensional scattering process is coherent in the third dimension, which in our case would be along the axis of the plume. A field of three-dimensional scattering would have many independent scatters distributed along the axis of symmetry. The ratio of the intensities of the two cases would be expected to vary as the number of independent samples and

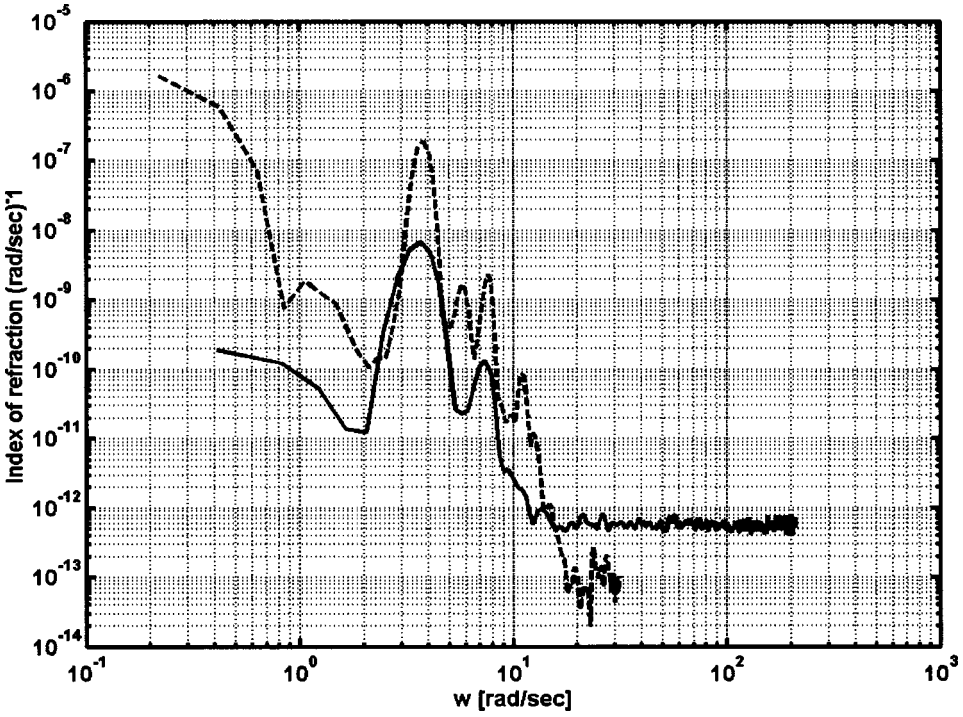


FIG. 5. One-dimensional spectrum of temperature for the unstable plume case. The dashed curve is calculated by using a thermometer with a correction for the high-frequency response included. The solid curve is the acoustically derived one-dimensional spectrum, calculated by integrating Fig. 4 over ''three-dimensional'' wave number space assuming isotropy.

would be of order K^*a, which agrees with the above analysis. Thus, a three-dimensional scattering process would require a higher level of variability in general than a two-dimensional one, since the two-dimensional process is by definition coherent in the third dimension. Equation (25) [and/or Eq. (26)] can be used for comparison with the one-dimensional spectrum estimated from the temperature probe measurements. The spectrum estimated from the temperature probe is corrected for the probe response time and converted to index of refraction. The one-dimensional spectral comparisons are shown in Fig. 5. The acoustically and mechanically (temperature probe) estimated spectra both contain peak values at 3.7 and 8 rad/s. The magnitude of the temperature probe-derived spectra is about an order of magnitude larger than the acoustically derived one. This is not unexpected, due to the acoustic estimate resulting from a bandlimited measurement and the assumption of isotropy for the unstable plume, which may not be valid. However, application of the isotropy assumption to scales of order the Bragg wave number (~mm wavelengths) might be expected to be a relatively good approximation, since some level of small-scale turbulence is present in the unstable plume case and turbulence at these scales is expected to be isotropic.[31] It is also interesting to note that it can be shown that the assumption of isotropy, analogous to the argument presented in contrasting two- and three-dimensional scattering, results in an upper bound estimate of the frequency spectrum $\Psi(\omega)$. That the acoustic estimate is below that from the temperature probe suggests that the explanation for the difference in these curves is that there are contributions to plume temporal variability from a range of spatial scales larger than that which can be resolved by the common Bragg wave number technique and would correspond to wave numbers (frequencies) below the resolution of the transducers being used.

B. Turbulent plume

Analysis similar to the unstable plume is conducted for the acoustic scattering from the turbulent plume. Figure 6(a) is a record of the three-dimensional wave number spectrum of the index of refraction fluctuations, $\Psi_{3D}(K,\tau)$, of the turbulent plume at $\tau=22.575$ s. This figure shows the results of common Bragg wave number configuration for the three-channel multistatic measurement. In the regions where there are common Bragg wave numbers between adjacent channels, the magnitude of the wave number spectrum between the channels is in good agreement but does contain regions of divergence between the channels. The pronounced structure that is present in the spectrum is typical for scattering from the turbulent plume. The nulls in the wave number spectrum can be as much as 30–40 dB below the mean spectral value. It should be noted that ping-to-ping records of the spectrum tend to remain highly coherent. The deep nulls indicate that the turbulence has a low Reynolds' number and thus is not fully developed. Figure 6(b) is a contour plot of the entire time series wave number spectrum of the index of refraction fluctuations for the turbulent plume. One of the first noticeable features in this figure, especially when compared to Fig. 3(b) of the unstable plume, is the absence of the periodic nature of the spectrum. As expected, the surface plot appears almost random in comparison, although there is still noticeable structure at smaller scales. Figure 6(b) also indicates the smoothly varying continuity over the entire bandwidth of wave numbers for the three-channel multistatic measurement. It is evident that the spectrum rolls off with increasing wave number, as expected for turbulence. The mean value of the three-dimensional wave number spectrum is given in Fig. 6(c). The solid black line represents the $-11/3$ power law expected for classical turbulence and is given for comparison with the acoustically estimated wave

number spectrum.[31] The acoustically estimated spectrum rolls off slightly faster than classical turbulence, not an unexpected result given the structure of the spectrum in Fig. 6(b) showing the nulls and the low value of the Reynolds number.

Figure 7 is a contour plot of the wave number/frequency spectrum for the data represented in Fig. 6(b), namely $\Psi_{3D}(K,\omega)$. Smoothly varying structure is noticeable throughout the entire wave number/frequency spectrum. There is also a perceptible trend in the spectrum such that the largest spectral values tend to coalesce to a confined region and form an apparent wave number-frequency dependence. A closer examination of these trends indicates that the slopes along the edges of the coalesced regions have nominal values between 0.5 and 3 cm/s. These values are well within the velocities expected for the turbulent plume, with maximum value of 3 cm/s associated with the mean vertical advection, and the lower value likely due to the turbulent eddies within the plume. Although not a part of the present work, coherent Doppler estimates from the acoustic scatter suggest the broadband averaged rms value for the fluid velocity of the present data set yield approximately 2 mm/s.

The one-dimensional spectral comparison of the index of refraction fluctuations in the turbulent plume between the acoustical and temperature probe estimates is given in Fig. 8. These calculations were performed in the same manner as

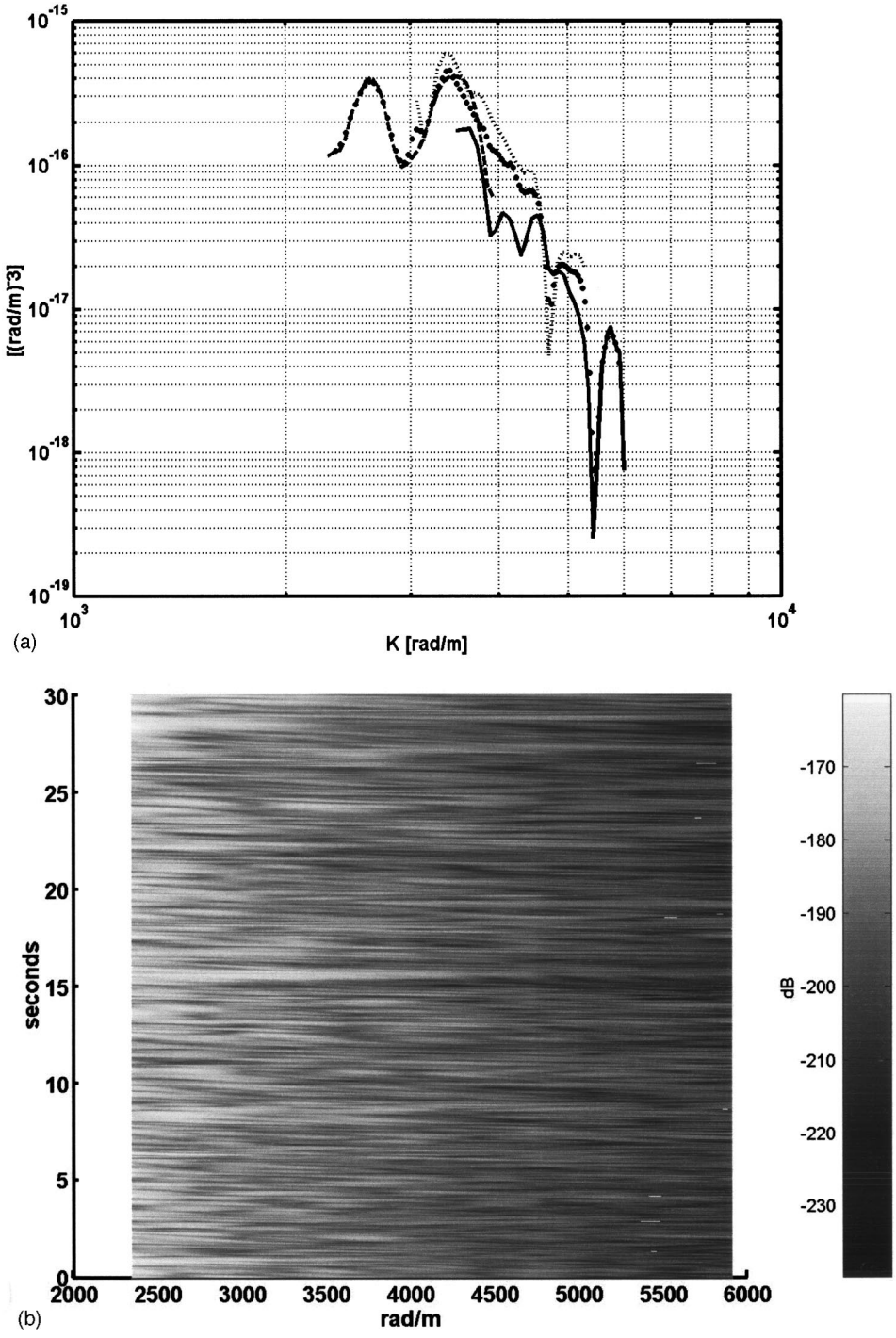
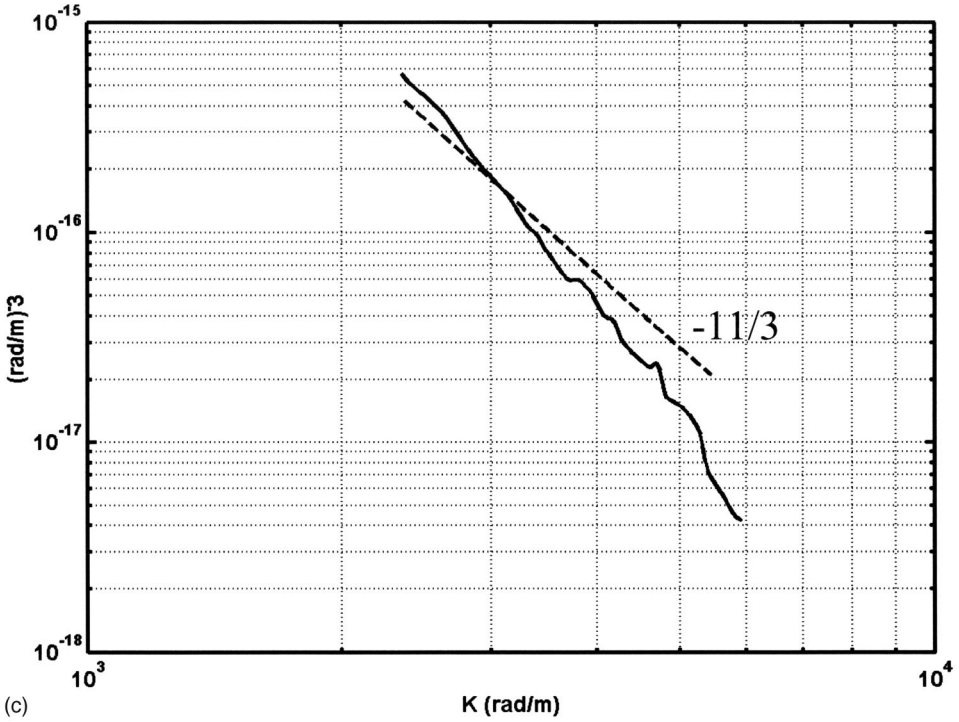


FIG. 6. (a) Common Bragg wave number spectrum of index of refraction fluctuations for the turbulent plume at $t=22.575$ s. It is formed by combining data from three separate receiver channels corresponding to scattering angles of 80 deg (dashed), 120 deg (dotted), 160 deg (solid). The black dots are an interpolation of the three curves over the entire range. (b) Time series of the interpolated common Bragg wave number spectrum of index of refraction fluctuations for the turbulent plume. It is formed by combining data from three separate receiver channels corresponding to scattering angles of 80 deg (green), 120 deg (blue), 160 deg (red). (c) Temporal mean wave number spectrum calculated from. The dashed line of $-11/3$ is the power law expected for an isotropic turbulent spectrum.



(c)

FIG. 6. (*Continued.*)

described for the unstable case. Between 20 and 50 rad/s there is remarkably good agreement between the two types of measurements, both in terms of magnitude and roll-off. Note that the acoustically estimated spectra levels off below 30 rad/s. Using an advection velocity for the plume of 1.5 cm/s yields a cutoff wave number of 2000 rad/m, which is of order of the lowest estimated wave number of 2300 rad/m, which can be estimated from the Bragg scattering geometry employed. If the acoustic signal were extended to lower wave numbers by either incorporating smaller scattering angles in the common Bragg wave number configuration or by using lower acoustic frequencies, the lower frequency part of the index of refraction (temperature) field as measured by the mechanical probe would be able to be estimated.

V. CONCLUSIONS

The issue addressed in this paper is the verification of application of far-field weak scattering theory to the case of acoustic scattering from thermally generated sound-speed variations in water. The motivation for the usage of high-

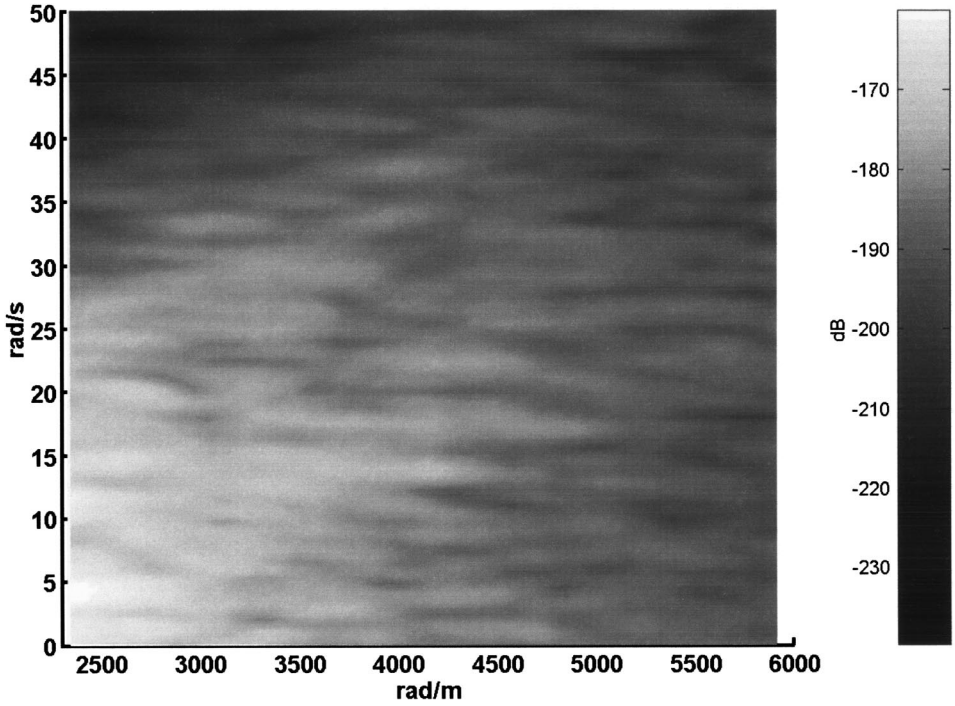


FIG. 7. Wave number/frequency spectrum of index of refraction fluctuations from a turbulent plume. Same data as those used in Fig. 6(b).

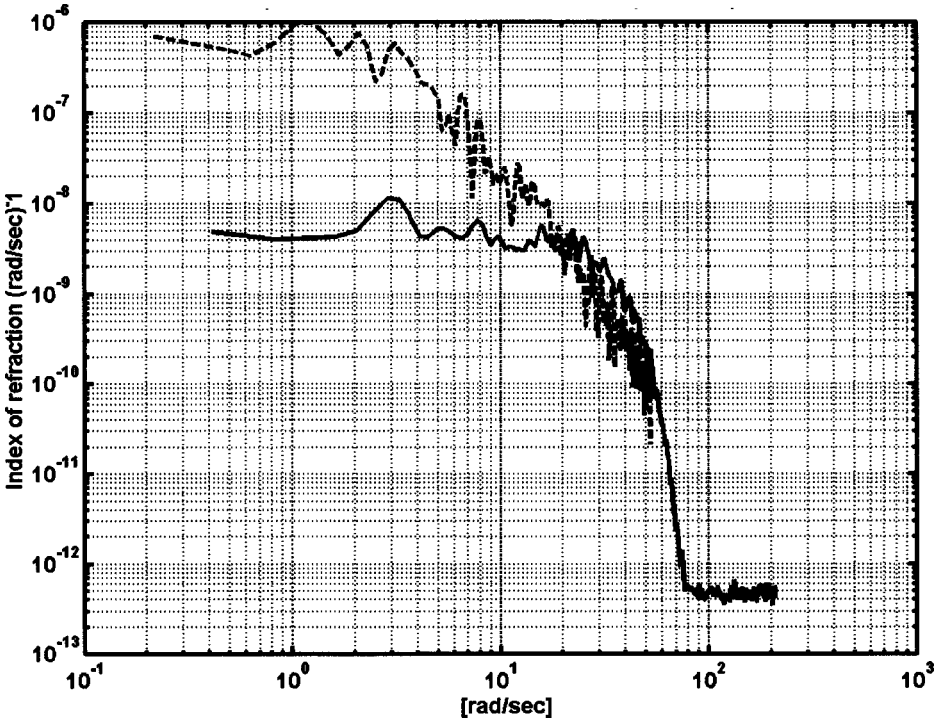


FIG. 8. One-dimensional spectrum of temperature for the turbulent plume case. The dashed curve is calculated by using a thermometer with a correction for the high-frequency response included. The solid curve is the acoustically derived one-dimensional spectrum, calculated by integrating Fig. 7 over ''three-dimensional'' wave number space assuming isotropy.

frequency broad-bandwidth multistatic acoustic scatter from medium variability is a consequence resulting from a prediction of far-field weak scattering theory, namely that the spectral magnitudes must be equivalent for the acoustic estimates from multistatic data taken coincidently in space and time for regions of overlapping Bragg wave numbers.

It has been shown that the multistatic acoustic scatter from both unstable and turbulent plumes validates the prediction made by the far-field weak scattering theory when ensemble averaging is employed to calculate the wave number spectrum. This is validated by the overlap of the spectral estimates shown in Figs. 3(a) and 6(a) for the unstable and buoyant plumes, respectively. The only *in situ* verification possible to date for our results has been to compare a reduced form of the spectral results, using either Eq. (23) or (24), to that obtained from a point temperature sensor. Note that the common Bragg wave number technique allows both a space and time measurement of the scattering field, whereas the temperature measurement only gives the time variability at a point.

We have also shown that unless some *a priori* information is known about the nature of the dimensionality of the scattering a two-dimensional inversion will yield the same shape wave number spectrum as a three-dimensional inversion. See Eqs. (6), (27), (28), and the discussion associated with those equations. The three-dimensional inversion appears to better describe the structure of both the unstable plume and turbulent cases. For the turbulent case, the acoustically estimated one-dimensional frequency spectrum was in very good agreement with that obtained by the temperature sensor over frequency ranges corresponding to the Bragg wave number; see Fig. 8.

The actual criterion of maximum anisotropy allowed using far-field Bragg scattering is not overly constraining. To see this, note as was discussed in Sec. II C the degree of anisotropy that can be tolerated given by Eq. (17b) is

$$2\theta_b \ll \alpha_z \sin\theta/2 \approx \alpha_z . \tag{29}$$

For our experiment, $\theta_b \cong 2.7/(ka)$, and corresponds to $0.05 \leqslant \theta_b \leqslant 0.08$ radians. Thus, the maximum degree of anisotropy which could be tolerated is expected to be of the order 0.1, and we conclude from our experimental results because of the common overlap of the spectral estimates shown in Figs. 3(a) and 6(a) that both the unstable and turbulent plume satisfy this criterion and do not have anisotropy values of θ_b less than of order 0.1.

ACKNOWLEDGMENTS

Acknowledgment is given to ONR 321 OA for support of this research and in addition for part of the contribution of John Oeschger to the Coastal Systems Station Independent Research Program.

APPENDIX:

The relative compressibility and density of a region of medium variability are given by $\gamma_\kappa = (\kappa - \kappa_o)/\kappa_o$, and $\gamma_\rho = (\rho - \rho_o)/\rho$, where the compressibility $\kappa = 1/\rho c^2$, where ρ and c are the anomalous density and sound speed, respectively.[21] Expressing these latter variables as the sum of the ambient value and its fluctuation, (i.e., $\rho = \rho_o + \rho'$ and $c = c_o + c'$), and since the fluctuations are expected to be small, the relative compressibility is approximated to first order by

$$\gamma_\kappa \cong -(2c'/c_o + \rho'/\rho_o). \tag{A1}$$

The relative density is approximated as

$$\gamma_\rho \cong \frac{\rho'}{\rho_o}. \tag{A2}$$

Given a nominal ambient water temperature of 20 °C and a maximum temperature difference from ambient of 5 °C, the terms in Eqs. (A1) and (A2) are readily determined[35] to be $2c'/c_o = 0.02$, and $\rho'/\rho = -0.001$. Thus, it is observed that thermally induced sound speed dominates density for the parameters given this paper. It should also be noted that effects due to salinity variability, while the subject of current investigation, are not pertinent in the present case of scattering from thermal driven variability in fresh water.

The relevant term for the relative fluid velocity is given by $2u'/c_o$.[1,20] For the case of the buoyant plume having a nominal vertical speed of 3 cm/s and used as the upper limit for the fluctuation of fluid velocity, $2u'/c_o = 0.000\,04$ and can be safely ignored compared to the thermally driven sound speed.

[1] L. Goodman, J. Oeschger, and D. Szargowicz, ''Ocean acoustics turbulence study: Acoustic scattering from a buoyant plume,'' J. Acoust. Soc. Am. **91**, 3212–3227 (1992).

[2] J. Oeschger and L. Goodman, ''Acoustic scattering from a thermally driven buoyant plume,'' J. Acoust. Soc. Am. **100**, 1451–1462 (1996).

[3] H. E. Seim, M. C. Gregg, and R. T. Miyamoto, ''Acoustic backscatter from turbulent microstructure,'' J. Atmos. Ocean. Technol. **12**, 367–380 (1995).

[4] P. H. Wiebe, T. K. Stanton, M. C. Benfield, D. G. Mountain, and C. H. Greene, ''High-frequency acoustic volume backscattering in the Georges Bank coastal region and its interpretation using scattering models,'' IEEE J. Ocean. Eng. **22**, 445–464 (1997).

[5] C. F. Greenlaw, ''Acoustical estimation of zooplankton populations,'' Limnol. Oceanogr. **24**, 226–242 (1979).

[6] D. V. Holliday, R. E. Pieper, and G. S. Kleppel, ''Determination of zooplankton size and distribution with multifrequency acoustic technology,'' J. Cons., Cons. Int. Explor. Mer **46**, 52–61 (1989).

[7] R. B. Mitson, Y. Simrad, and C. Goss, ''Use of two-frequency algorithm to determine size and abundance of plankton in three widely spaced locations,'' ICES J. Mar. Sci. **53**, 209–215 (1996).

[8] T. K. Stanton, ''Sound scattering by cylinders of finite length. I. Fluid cylinders,'' J. Acoust. Soc. Am. **83**, 55–63 (1988).

[9] T. K. Stanton, ''Sound scattering by cylinders of finite length. III. Deformed cylinders,'' J. Acoust. Soc. Am. **86**, 691–705 (1989).

[10] Z. Ye, ''A novel approach to sound scattering by cylinders of finite length,'' J. Acoust. Soc. Am. **102**, 877–884 (1997).

[11] D. V. Holliday, ''Extracting bio-physical information from the acoustic signitures of marine organizms,'' in *Ocean Sound Scattering Prediction* (Plenum, New York, 1977).

[12] C. F. Greenlaw and R. K. Johnson, ''Multiple-frequency acoustical estimation,'' Biol. Oceanogr. **2**, 227–252 (1983).

[13] R. C. Waag, D. Dalecki, and W. A. Smith, ''Estimates of wavefront distortion from measurements of scattering by model random media and calf liver,'' J. Acoust. Soc. Am. **85**, 406–415 (1989).

[14] R. C. Waag, D. Dalecki, and P. E. Christopher, ''Spectral power determination of compressibility and density variations in model media and calf liver using ultrasound,'' J. Acoust. Soc. Am. **85**, 423–431 (1989).

[15] R. C. Waag, J. P. Astheimer, and J. F. Smith III, ''Analysis and computations of measurements system effects in ultrasonic scattering experiments,'' J. Acoust. Soc. Am. **91**, 1284–1297 (1992).

[16] P. D. Thorne, P. J. Hardcastle, and R. L. Soulsby, ''Analysis of acoustic measurements of suspended sediments,'' J. Geophys. Res. **98**, 899–910 (1993).

[17] J. J. Tsai, H. Huang, and P. R. Proni, ''Acoustic observations of bottom surge from dredged material discharge in the open ocean,'' Chem. Ecol. **10**, 71–85 (1995).

[18] P. D. Thorne and P. J. Hardcastle, ''Acoustic measurements of suspended sediments in turbulent currents and comparison with *in situ* samples,'' J. Acoust. Soc. Am. **101**, 2603–2614 (1997).

[19] A. M. Crawford and A. E. Hay, ''Determining suspended sand size and concentration from multifrequency acoustic backscatter,'' J. Acoust. Soc. Am. **94**, 3312–3324 (1993).

[20] A. E. Hay and J. Sheng, ''Vertical profiles of suspended sand concentration and size from multifrequency acoustic backscatter,'' J. Geophys. Res. **97**, 15661–15677 (1992).

[21] T. K. Stanton, P. H. Wiebe, D. Chu, and L. Goodman, ''Acoustic characterization and discrimination of marine zooplankton and turbulence,'' ICES J. Mar. Sci. **51**, 469–479 (1994).

[22] S. A. Thorpe and J. M. Brubaker, ''Observations of sound reflection by temperature microstructure,'' Limnol. Oceanogr. **28**, 601–613 (1983).

[23] L. Goodman, ''Acoustic scattering from ocean microstructure,'' J. Geophys. Res. **95**, 11557–11573 (1990).

[24] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, Princeton, NJ, 1968).

[25] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978), Vol. 1, Chap. 2.

[26] V. I. Tatarski, *Wave Propagation in a Turbulent Medium* (McGraw-Hill, New York, 1961).

[27] J. Oeschger, ''Acoustic scattering from a laminar plume,'' Ph.D. thesis, University of Rhode Island, 1993.

[28] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982).

[29] F. A. Jenkins and H. E. White, *Fundamentals of Optics* (McGraw-Hill, New York, 1976).

[30] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978), Vol. 2, Chap. 16.

[31] G. K. Batchelor, *Homogeneous Turbulence* (Cambridge University Press, New York, 1967).

[32] J. Andrzej Domaradzki and Wei Liu, ''An analysis of subgrid-scale interactions in numerically simulated isotropic turbulence,'' Phys. Fluids A **5**, 1747–1759 (1993).

[33] R. S. Ragallo and P. Moin, ''Numerical simulation of turbulent flows,'' Annu. Rev. Fluid Mech. **16**, 99 (1984).

[34] A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics* (MIT Press, Cambridge, MA, 1979).

[35] P. Fofonoff and R. C. Millard, Jr., ''Algorithms for computation of fundamental properties of seawater,'' Unesco Tech. Papers in Mar. Sci., No. 44, 1983.

Modeling signal loss in surficial marine sediments containing occluded gas

Trevor Gardner[a)]
Department of Engineering Science, University of Oxford, Parks Road Oxford OX2 7LD, England and School of Sport and Exercise Sciences, University of Birmingham, Edgbaston, Birmingham BTT 2TT, England

(Received 4 November 2001; revised 28 October 2002; accepted 4 December 2002)

The presence of occluded gas in inland lakes, harbor muds, and surficial marine sediments is well documented. Surficial gassy sediments cause underlying beds to be acoustically impenetrable to seismic surveys; therefore, the modeling of signal loss arising from mudline reflection and transmission absorption is of particular interest. The Anderson and Hampton [J. Acoust. Soc. Am. **67**, 1890–1903 (1980)] model for attenuation in gassy sediments was evaluated against the physical and acoustical properties of eight laboratory silty clay soils containing different amounts of occluded gas in bubbles of 0.2- to 1.8-mm diameter. The model was shown to give good agreement with measured data over the lower frequencies of bubble resonance and above resonance. It did not agree with measured data at frequencies below resonance, for which the model did not simulate the bulk properties of the gassy soils. The Mackenzie [J. Acoust. Soc. Am. **32**, 221–231 (1960)] model for reflection loss was also examined for the gassy soils. The maximum reflection losses of 6 dB, at a grazing angle of 40°, does not wholly support speculation by Levin [Geophysics **27**, 35–47 (1962)] of highly reflective pressure-release boundaries arising from substantial reflection and absorption losses in gassy sediments. It was found that mudlines formed from sediments with significant occluded gas may be successfully penetrated, although the substantial absorption loss arising from signal transmission through the sediment prevents penetration of the surficial layers to much beyond a meter in depth.
[DOI: 10.1121/1.1542731]

PACS numbers: 43.30.Ma, 43.30.Pc, 43.30.Vh [DLB]

I. INTRODUCTION

The presence of occluded gas in harbor muds and in surficial marine sediments is well documented.[1–3] It is commonly found in loosely compacted silty clay sediments such as lake beds, river deltas, and harbors where tributary streams bring a supply of organic matter that settles out with the silt and clay in comparatively slow moving, deeper waters. Free ammonia, hydrogen sulfide, or methane gas arising from anaerobic decomposition may exist either as bubbles within fluid-filled interparticle voids, or as gas pockets formed from the solid–liquid matrix.[4] Gassy sediments are often detected by seismic surveys during mapping or exploration, because signal loss due to scatter, reflection, and absorption may cause these sediments to become acoustically impenetrable.[5–7] Although this has initiated some experimental testing and theoretical modeling to determine the acoustic properties of gas bearing sediments,[8–11] little verifiable progress has been made in predicting signal losses arising from reflection and attenuation.

Lyons *et al.*[12] modeled acoustic signals returned from the gassy sediments of Eckernfoerde Bay, Germany, and showed sediment scatter to be dominated by the presence of methane gas bubbles. Gardner and Sills[13] showed that compressional wave sound speed in laboratory gassy soils characterized by Gardner[14] could be predicted using measurements of the soil's physical properties and estimates of the soil's shear and Young's moduli. Estimates were based upon the studies of Wheeler[15–17] on gassy soils at high mechanical strain and of Jardine *et al.*[18] on saturated soil at low strain. The soil's physical properties were then used by the acoustic model developed by Anderson and Hampton[8] to predict sound speeds for gassy soils. The expression for attenuation derived by Anderson and Hampton[8] is further developed from Silberman's[19] expression for gas bubbles in a water medium by including the term $4G/3$ to provide the medium with the shear rigidity (G) of a sediment. Although they carried out tests on gas-filled plastic bubbles to verify the effects of bubble resonance, the performance of their attenuation model has not been investigated to date.

Verified models for reflection losses for non-normal incidence at submerged gasified mudlines are so far unavailable, although predictions of normal incident reflection loss at interfaces of gas bubbles in water give good correlation with experimental data.[20] These indicate substantial losses of 16 dB at frequencies of peak bubble resonance. Some progress in modeling non-normal incidence has been made by Mackenzie,[21] who further developed Lord Rayleigh's[22] equation to predict the reflection coefficient for a signal normal to a liquid/liquid interface to accommodate varying signal grazing angles (between the incident signal and the interface). Sediment absorption loss was accommodated by defining the wave number for the sediment as complex to extend the model to a liquid/sediment interface. Mackenzie's

a)Electronic mail: t.gardner@bham.ac.uk

model[21] has, so far, only been validated for nongassy sediments. This has been achieved by the *in situ* measurements of Mackenzie,[21] at signal grazing angles of 12° and 90° (at 4 to 16 kHz), and by Muir *et al.*[23] at angles of 14° to 78° (at 20 kHz).

It is believed that gas-induced influences upon signal absorption that arise from changes in the bulk material properties of the soil at the interface can be embodied in the Mackenzie model for calculating reflection loss. Although this model is currently unable to predict specular reflection at submerged mudlines over frequencies initiating bubble resonance, it should predict reflection losses for gassy soils at frequencies below resonance. This is so, since at these lower frequencies it has been shown that the sediment responds acoustically as a single phase medium characterized by its bulk (combined gas–sediment) elastic properties.[8,14] Thus, if Mackenzie's expression is used to predict reflection loss using acoustic properties relevant to the response of gassy soils at frequencies below resonance, solutions must be valid within this signal range. In the case of gasified sediments, it has not been possible previously to validate the Anderson and Hampton model for attenuation or to use the Mackenzie model to predict reflection loss, since the comprehensive structural and material properties required to fully characterize a gasified sediment have been unavailable. This is no longer the case, since Gardner[14] recently fully defined the parameters required by each model to simulate the acoustic behavior of eight gassy soils.

Therefore, the present study investigates whether mudline reflection and transmission absorption loss in gassy soil may be predicted successfully for the Gardner soils using current theory. Models developed by Anderson and Hampton[8] for attenuation and by Mackenzie *et al.*[21] for reflection were used to predict attenuation and absorption in the Gardner gassy soils. Predictions were based upon the soils' physical properties measured by Gardner,[14] and their elastic properties estimated by Gardner and Sills.[13] Predicted attenuation was evaluated against measured data reported by Gardner,[14] but similar data on measured reflection losses were unavailable for comparison with predicted reflection losses. Mackenzie's model, however, has been validated for saturated sediments, in which bulk elastic properties determine the soil's physical and acoustical behavior. It is therefore believed that the validation also extends to gassy sediments at frequencies below bubble resonance, since at these frequencies it is again the bulk elastic properties of the soils that determine their physical and acoustical behavior.[8,14] Reflection losses predicted by the present study are therefore considered useful in examining the performance of seismic surveys over gasified mudlines.

II. METHOD

A. Measured attenuation losses by Gardner (Ref. 15)

The physical properties of the eight laboratory soils, each of different gas content, reported by Gardner[14] are shown in Tables I and II. These gassy soils were produced with uniformly distributed spherical bubbles by releasing methane gas into estuarine silty clay under controlled conditions that simulated the natural process of gas-bubble formation in sediment. Twenty-five kPa of total stress was applied to the soil to model the consolidation of a surficial sediment (1–2 m in depth) in relatively shallow water. Table I shows soil gas fractions (the proportion of methane gas volume to total soil volume) from 0.003 57 to 0.198, which encompass the range of gas contents found in naturally occurring gassy soils. Table II shows the range of bubble diameters contained in each soil. Acoustic compressional waves were propagated through the soil and collected by piezoelectric transmitting and receiving transducers at the upper and lower soil surfaces. These were examined by Fourier analysis. This was repeated after replacing each soil sample with a liquid of the same acoustic impedance and signal path length. Attenuation was found using Eq. (1)

$$\alpha = \frac{10}{L}\log_{10}\left[\frac{V_s}{V_l}\right]^2, \tag{1}$$

where V_s and V_l are the voltage amplitudes of the soil and liquid, and L is the signal path length for the soil and the liquid. Absorption losses in the liquid were assumed negligible in comparison with the soil, and transducer transmission, transducer interface, and radiation losses through the medium were neglected since in both mediums losses were equal and were negated by Eq. (1). Attenuations of the gassy soils measured by Gardner[14] are shown in Fig. 1.

TABLE I. Physical properties of the gassy soils.

Soil	Gas fraction	Degree of saturation (%)	Bulk density (kg/m^3)	Void ratio
50-1	0.198	72	1322	2.311
50-2	0.173	75	1363	2.215
52-1	0.082	87	1513	1.899
52-2	0.115	83	1459	2.003
53-1	0.005 11	99.18	1639	1.674
53-2	0.004 07	99.35	1640	1.674
51-1	0.004 84	99.23	1640	1.674
51-2	0.003 57	99.43	1643	1.667

B. Predicting attenuation using the Anderson and Hampton model (Ref. 1)

Attenuations for the Gardner soils over the frequency range influenced by bubble resonances were predicted in the present study using the Anderson and Hampton model[8]

TABLE II. Diameters and resonance frequencies of bubble groups within the gassy soils.

	Bubble diameter (μm)		Resonance frequency (kHz) for	
Soil	Minimum size	Maximum size	Minimum bubble size	Maximum bubble size
50-1	229	1746	793	104
50-2	247	1746	734	104
52-1	276	807	653	223
52-2	348	1605	516	112
53-1	137	458	1302	391
53-2	137	276	1297	649
51-1	137	385	1297	463
51-2	174	669	1039	270

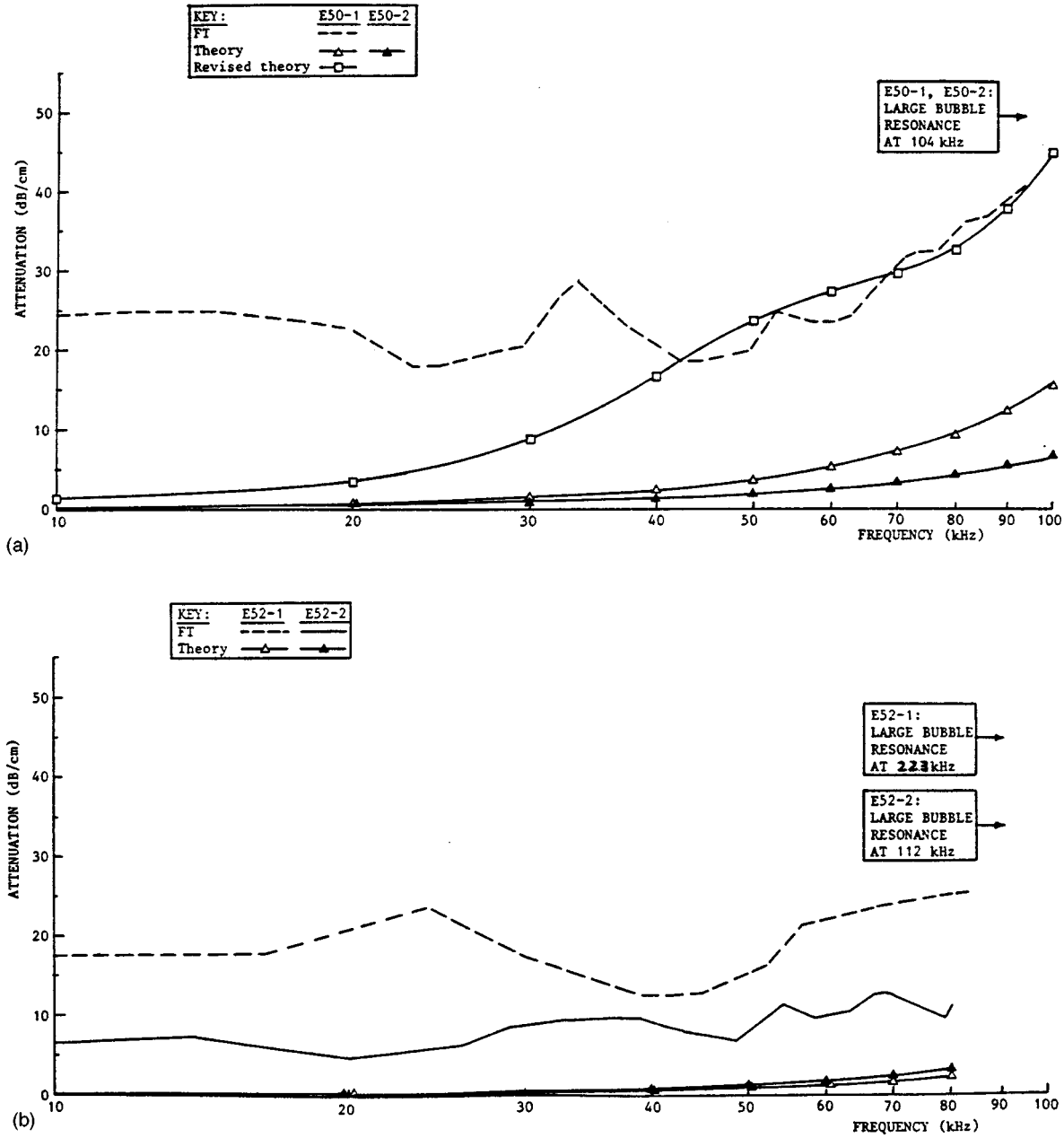


FIG. 1. Attenuations of their eight gassy soils measured by Gardner (Ref. 8) (marked FT in the key). For direct comparison, attenuations of the same soils are predicted by the Anderson and Hampton model, using values of K_{sat} and G_{sat} based on case (i) (marked theory in the key). (Note that symbols do not denote data points; actual points are at 2.441-kHz intervals.)

$$\alpha = \frac{\pi f}{c_o} \cdot \frac{c}{c_o} \cdot \frac{KY_1}{\gamma P_o + \frac{4G}{3}}. \tag{2}$$

Parameters are identified in Appendix A, where values are provided from the physical and acoustic properties of the Gardner soils. The volume concentration of gas in each bubble size provided by Gardner[14] is accommodated within the Anderson and Hampton[8] constants X_1 and Y_1. The dilatational modulus of the (solid–fluid) saturated soil matrix (K_{sat}) is used in the Anderson and Hampton acoustic equation, rather than the bulk modulus of the gassy soil (K). This is because their equation was developed from Silberman,[9] where K was defined as the modulus of the host component in a biphasic gassy medium. The same argument applies for the matrix shear rigidity (G_{sat}), which was used instead of G. In view of acoustic predictions being extremely sensitive to these particular material properties,[13] alternative approaches were used to calculate K_{sat} and G_{sat} (the dilatational and shear moduli for saturated media). This ensured that Anderson and Hampton's model[8] for attenuation was evaluated using the best input data available. Case (i) adopts the method prescribed by Anderson and Hampton to estimate K_{sat} and G_{sat}, for their Eq. (2), and case (ii) describes the method initially used by Gardner and Sills[13] specifically for the Gardner[14] soils. Case (iii) adopts refined values of K_{sat} and G_{sat} shown by Gardner and Sills to improve the agreement between the sound speeds measured by Gardner[14] and those predicted by Anderson and Hampton's model.

1. Case i—upper bound estimates prescribed by Anderson and Hampton (Ref. 8)

Dynamic measurements on saturated mediums were used to provide upper bound values of sound speed. G_{sat} (0.13 GPa) was estimated from dynamic measurements on saturated surficial sediments by Smith[24] and Hamilton,[25] and the corresponding K_{sat} (3.65 GPa) was calculated using expressions for moduli developed by Gassman.[26]

2. Case ii—lower bound estimates taken from Jardine et al. (Ref. 18), Wheeler (Refs. 15–17)

G_{sat} (0.05 GPa) was estimated from the high-rate, load-displacement soil tests of Jardine *et al.*[18] at strains of 0.01%, since this was appropriate to acoustic signals of strain 0.01% to 0.001%. The value of K_{sat} (4.211 GPa) that corresponded with this value of G_{sat} was obtained from the experimental work of Wheeler[15–17] on saturated soils prepared from the same soil batch and in the same manner as the Gardner large-bubble soils. This procedure is justified, since it is evident from Wheeler and Gardner[27] that bulk properties of the Gardner soils[14] estimated from these sources produce sound speeds that agree well with the measured data at frequencies below bubble resonance.

3. Case iii—Refined estimates

Gardner and Sills[13] more accurately simulated the mechanical behavior of the Gardner soils by refining the values of K_{sat}, G_{sat}, and Q to improve the fit between sound speeds measured by Gardner,[14] and those predicted by the Anderson and Hampton model. Q is the specific dissipation function (the inverse of the damping constant defined by Anderson and Hampton[8]). To similarly improve the fit between predicted and measured values of attenuation, Gardner and Sill's refined values of K_{sat} (3570 MPa), G_{sat} (20 MPa), and Q (4) were used in Eq. (2).

C. Resonance frequency predicted by the Anderson and Hampton model (Ref. 8)

The frequencies at which the largest bubbles in each soil are expected to reach peak resonance are calculated from Anderson and Hampton's[8] expression

$$f_o = \frac{1}{2\pi r}\sqrt{\frac{3\gamma P_o}{A\rho_s} + \frac{4G_{sat}}{\rho_s}}. \tag{3}$$

All parameters are defined in Appendix A, and the largest bubble sizes for each soil are shown in Table II.

D. Predicting reflection using the Mackenzie model (Ref. 2)

The method used by Mackenzie[21] to determine reflection loss (L_R) is described in Appendix B.

$$L_R = 10\log_{10}\left[(1+R)^2\frac{\rho_2 c_2}{\rho_1 c_1}\right]. \tag{4}$$

Equation (4) was used to predict reflection coefficients at different interface grazing angles for five of the Gardner soils covering the range from a saturated soil (containing no gas) to a soil of high gas content. Bulk density, compression wave sound speed and attenuation were taken from the properties of the soils given in Appendix A and Gardner.[14] Predicted reflection coefficients of the saturated soil and the soils of low and high gas content (E53-1 and E50-1, respectively), were then used to predict reflection loss. Sound speed (1532.3 m/s) and bulk density (1025 kg/m^3) of the salt water were taken from the velocity ratios of Table I and the densities of Table II, Hamilton,[25] since they are relevant to average continental shelf conditions overlying silty clay sediments. Saturated sediment properties were taken as 1502.5 m/s (measured) and 1648 kg/m^3 [Gardner and Sills[13] Eq. (A2), Appendix A]. Comparative reflection losses were also predicted by Mackenzie's model[21] based on physical properties of saturated sediments reported by Hamilton.[25,28] This assisted in assessing the influence of gas on reflection loss in silty clays to course sands.

III. RESULTS

A. Attenuation

Figure 2 shows attenuations of the eight Gardner soils, predicted by Eq. (2) (Anderson and Hampton[8]), based on the different values of G_{sat} and K_{sat}. Case (i) uses the upper bound K_{sat} and G_{sat} defined by Anderson and Hampton, and case (ii) uses the lower bound K_{sat} and G_{sat} used by Gardner and Sills.[13] Peak attenuations at frequencies of peak resonance seem unrealistically high (100 to 400 dB/cm) in both cases for the soils of high gas content (E50 and E52). Comparisons between predicted and measured data are also made with other studies. Figure 3 comprises the few acoustic measurements on gassy soils from the literature,[11,29,30] of largely indeterminate gas content. The highest losses of between 26 and 75 dB/cm are reported by Nyborg *et al.*[11] and these embrace the losses measured by Gardner[14] which occurred at signal frequencies of 10 to 35 kHz. Over nondispersive frequencies above bubble resonance, attenuations predicted for the Gardner soils (1 to 10 dB/cm) accurately simulate the response of saturated silty sands containing no gas (0.006, at 10 kHz—13 dB/cm, at 388 kHz, Hamilton[31]).

Figure 1 shows attenuation data predicted by case (i) over the frequency range for which attenuation was measured by Gardner.[14] The predicted curves confirm that Gardner's measured data extend over a narrow frequency range below bubble resonance and the lower frequencies that initiate bubble resonances. However, there is little correlation between predicted and measured data in terms of their magnitude and curve shape. Despite the irregular pattern of the curve for the measured data, predicted loss underestimates measured loss for each soil (by around 20 dB/cm), indicating that the moduli and damping properties of the soils simulated by case (i) (Anderson and Hampton) may be inaccurate. This is supported by evidence from Fig. 1(a), that attenuations based upon the refined properties Q, G, and K of case (iii) improve the correlation between measured and predicted data at frequencies above 40 kHz. At this frequency, bubble resonances begin to elevate the level of attenuation. For the remaining soils, the refined soil properties will similarly improve the correlation with measured data over the same fre-

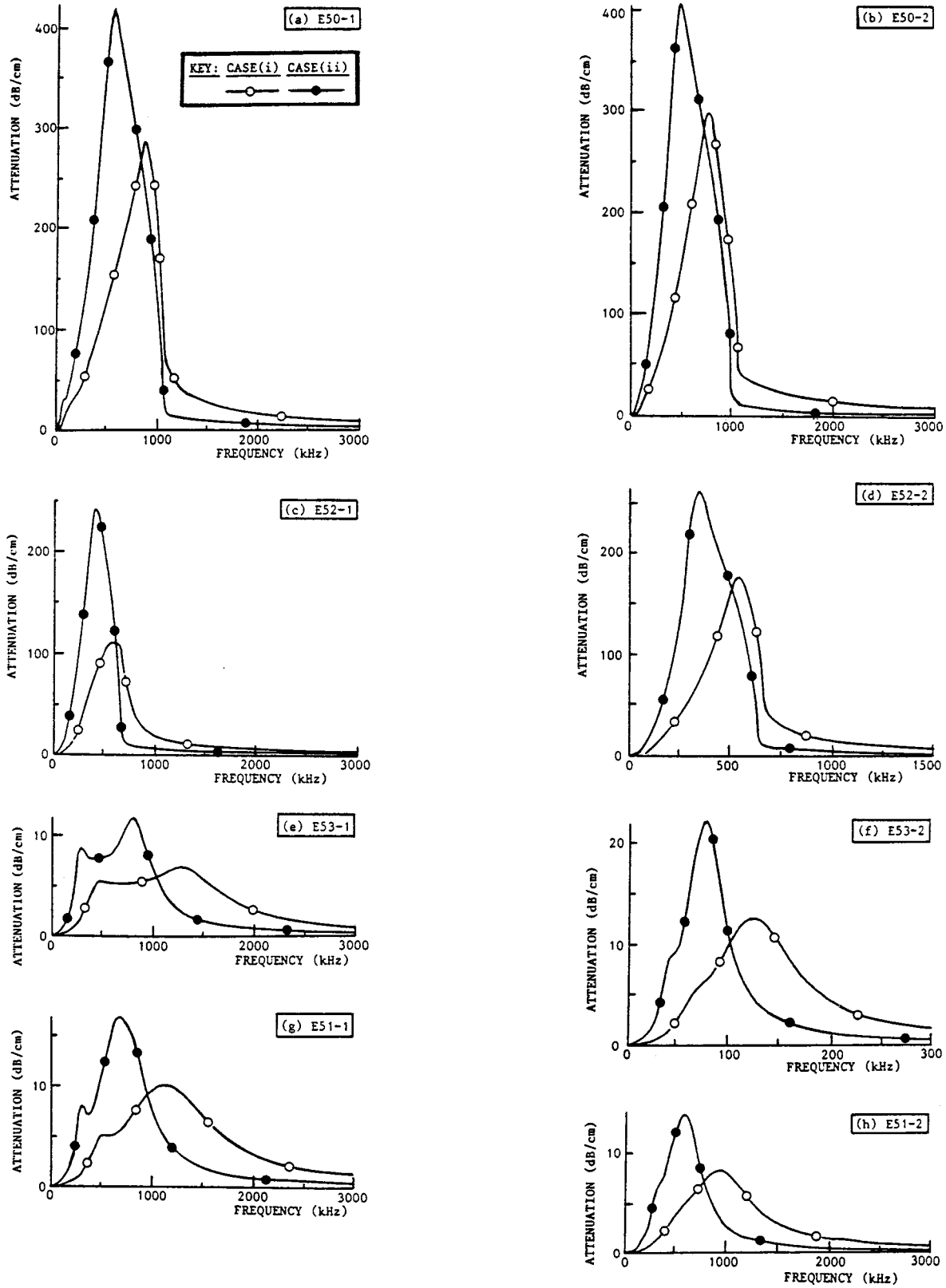


FIG. 2. Attenuations for the eight Gardner soils calculated by the Anderson and Hampton model. The model uses the dilatational and shear moduli of saturated soil (K_{sat} and G_{sat}) based on: case (i) Anderson and Hampton (Ref. 8), and case (ii) Jardine *et al.* (Ref. 18) and Wheeler (Refs. 15–17). (Note that symbols do not denote data points; actual points are at 2.441-kHz intervals.)

quencies. This is so, since attenuation is again consistently underestimated by around 20 dB/cm. There is no similar improvement of fit for attenuations of Fig. 1(a) between 10 and 40 kHz; here, predicted values remain around 25 dB/cm lower than measured values.

Marked on each of Figs. 1(a)–(d) as bars is the approximate frequencies at which the largest bubbles in each soil are expected to reach peak resonance [Table II calculated by Eq. (3)]. The positioning of the bars indicates the resonance frequencies predicted for each of the larger bubble sizes (maxi-

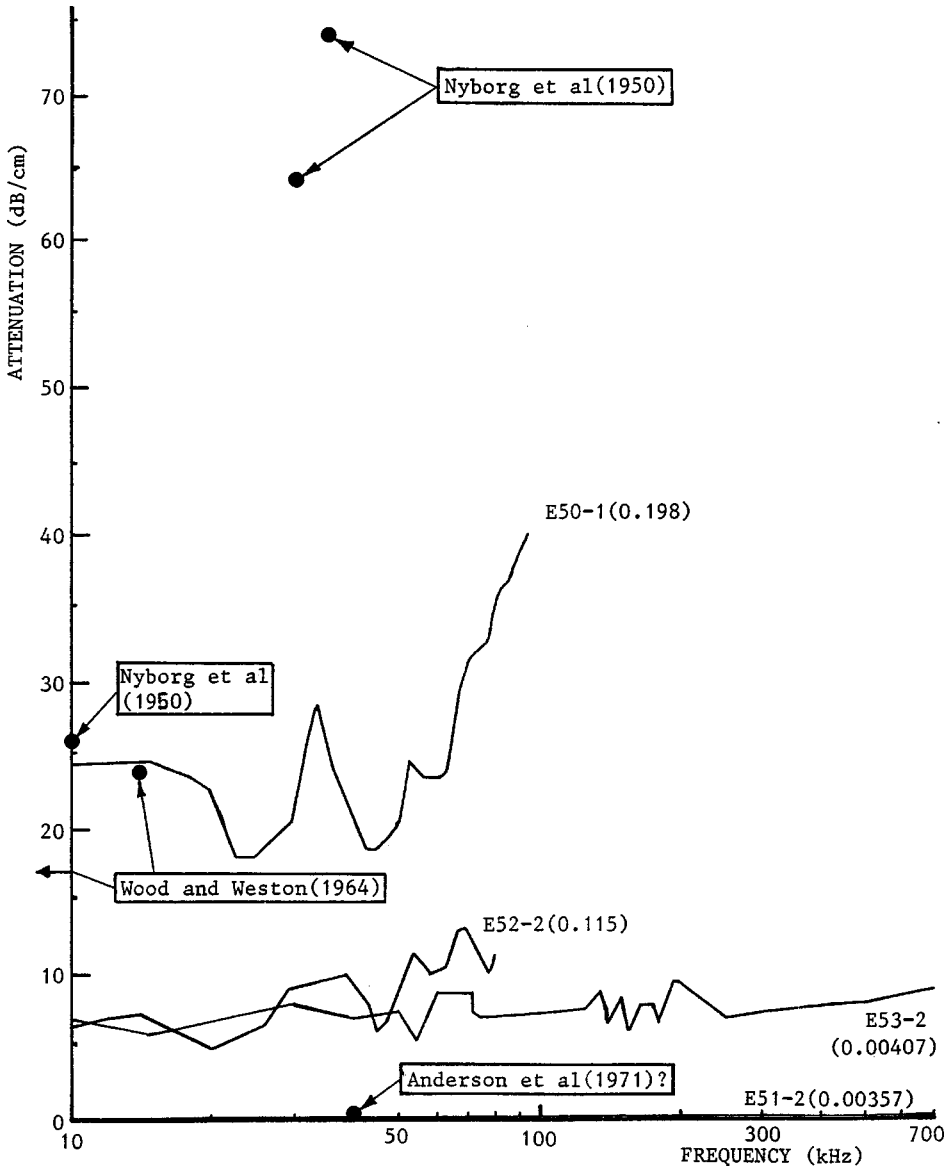


FIG. 3. Attenuations measured in gassy soils reported in the literature. Currently, only the Gardner (Ref. 14) gassy soils (prefixed "E," with gas fractions) are characterized fully (physically and acoustically). Other studies do not provide reliable gas contents, bubble sizes, and mechanical parameters, and the relationships between the signal frequencies transmitted, and those initiating bubble resonance effects, are unknown.

mum diameter on the left, reducing in diameter towards the right), and the relative heights of the bars indicate the relative gas fractions contained in each bubble size. Resonance frequencies are 104 to 223 kHz for the soils of high gas content (E50, E52) and 350 to 600 kHz for the soils of lower gas content (E51, E53). It can be seen that resonances occur at frequencies consistently above the upturn in the predicted attenuation curves of Fig. 1 that denote the onset of resonance effects.

B. Reflection

Reflection coefficients at different grazing angles are shown in Fig. 4. The soil with almost no gas (E53-2) exhibits a similar sign reversal to the saturated soil, whereas E53-1 containing slightly more gas exhibits a very different trend of variation characteristic of soils of higher gas content (E52-1 and E50-1). The influence of only a small amount of gas upon the soil's interface reflection properties arises from its effect on sediment sound speed and attenuation.[14] In Fig. 5, loss for the slightly gassy soil (E53-1) is significant only for angles of less than 40°, whereas for the very gassy soil (E50-1) loss is greater than 2.5 dB at 90° and rises to 6 dB at 40°; this is the angular range over which most seismic profiling devices operate.

For the saturated sediments shown in Fig. 6, the absence of a critical angle, below which all energy is reflected, is typical of a silty clay for which the sound speed is generally less than in the overlying water. This is generally not the case for more granular soils, such as course sands and gravels. At 90°, the nongassy silty clay exhibits a loss of only 0.15 dB, which is comparable to the loss of around 0.25 dB predicted by the Mackenzie model for the Gardner saturated soil. The silty clay gassy soils of Gardner exhibit a loss of 3 dB. Additionally, the 3-dB range of variation for gassy soils of different gas content is much greater than the 0.6-dB range of variation for nongassy soils of different particle classification (silty clay to coarse sand).

IV. DISCUSSION

Attenuations reported in the literature are consistently lower than peak attenuations predicted by the model. There are several explanations for this. Nyborg *et al.*'s[11] soils were

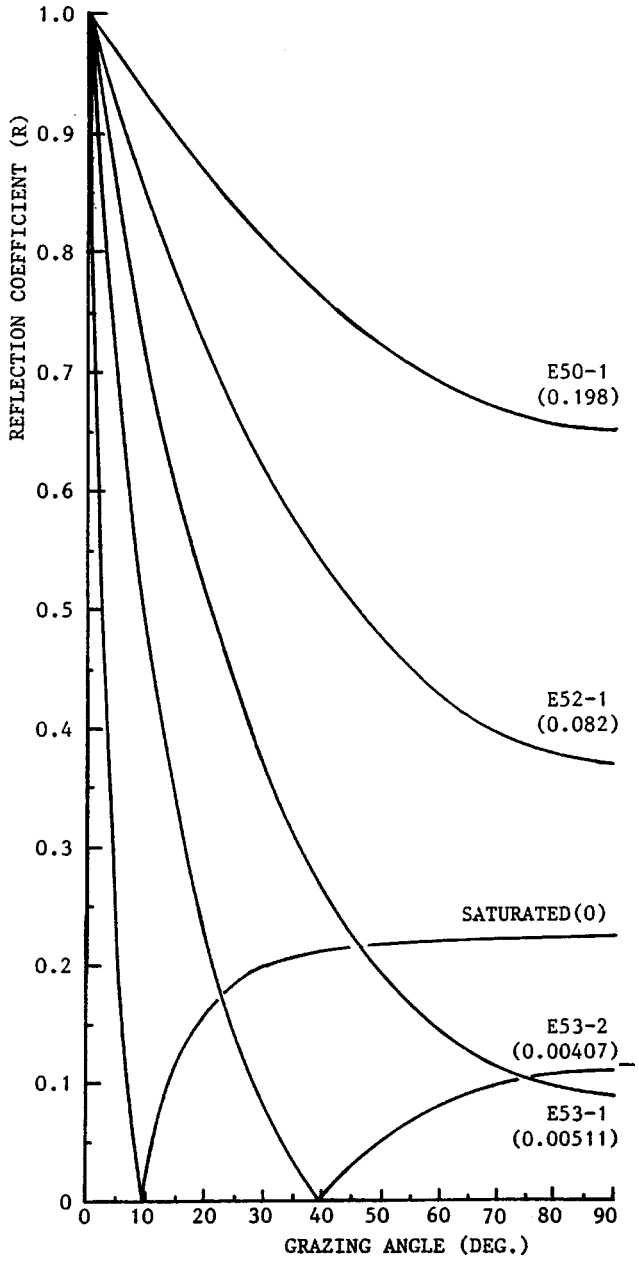


FIG. 4. Reflection angles for the Gardner (Ref. 14) soils calculated by the Mackenzie (Ref. 21) model. The soils of significant gas content (gas fraction >0.005) do not exhibit the sign reversal of normal saturated soils.

loose Hagerstown silty loam gasified by entraining air. Though they were of similar physical properties to the Gardner soils, in terms of particle type and size, their consolidation may have been lighter and the method of gasification suggests gas fractions were in excess of 0.2. In this case, the bulk moduli may be much lower than in the Gardner soils, leading to the lower reported attenuations at frequencies of resonance. This may also be the case for the unconsolidated muds of Wood and Weston[29] that contain entrained gas, and those of Anderson *et al.* from Lake Austin.[30] Another explanation for the difference may be that measured data are consistently at frequencies below resonance at which losses are much reduced.

As shown with sound speed,[13] the Anderson and Hampton model for attenuation is sensitive to inaccurate estimates

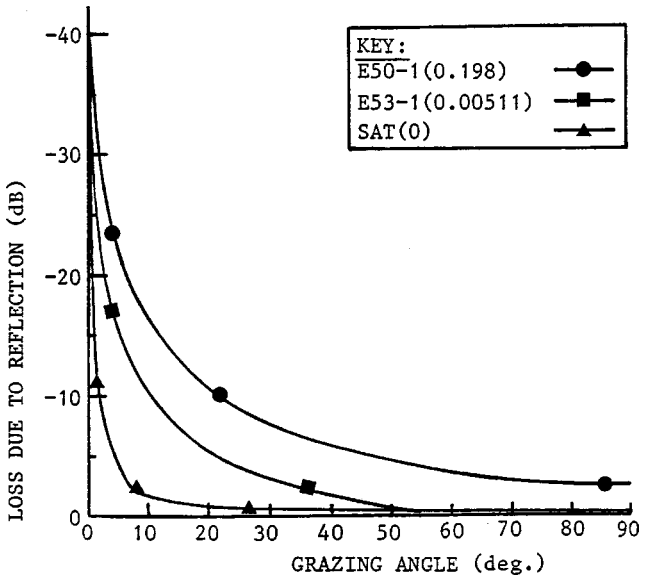


FIG. 5. Reflection losses for the Gardner (Ref. 14) soils. Losses for the Gardner gassy soils (of gas fraction >0.005) are very significant at signal interface grazing angles larger than 40°. (Note that symbols do not denote data points; actual points are at degree intervals.)

of the soil's material properties G_{sat}, K_{sat}, and Q. However, the refined soil properties prescribed by Gardner and Sills[13] indicate that the model can be accurate at least over the lowest frequencies influenced by bubble resonance. Also, predictions for frequencies above those influenced by bubble resonance are comparable with attenuation data from the literature on nongasified sediments. This supports Anderson and Hampton's theory that, at these frequencies, gassy soils respond as saturated sediments. This is so, since $\frac{1}{4}$-wavelengths are smaller than gas-bubble diameters; therefore, the acoustic response of the soil is dominated by the saturated soil matrix comprising the pore fluid and solid particles.[14] For frequencies of resonance, the unrealistically high peaks predicted in Fig. 2 and the reduced disparity between predicted attenuation and measured data arising from increased damping, further validates the properties of the Gardner soils that were refined by Gardner and Sills.[13] Gardner and Sill's demonstrated that the acoustical behavior of the Gardner soft soils is best simulated by Anderson and Hampton's model using a specific dissipation function (Q) of 4, rather than by determining Q from the model.

It is worth noting that the frequency above which bubble resonance effects elevate attenuations predicted by the model is consistently lower than the frequency at which the largest bubbles in each soil are predicted by the model to achieve peak resonance. Anderson and Hampton point out, for single-sized gas bubbles in water, dispersion is not closely confined to the frequency of peak resonance. It occurs over an extended range of frequencies indicated by $10^{-2}<Kr<10^{0}$ (where K and r are the wave number and bubble radius). Therefore, in the Gardner soils, the extended range of frequencies influenced by dispersion arises from both the numerous bubble sizes present in the soil (137 to 1746 μm, Table II) and the range of frequencies at which each bubble size resonates (1297 to 104 kHz, Table II of f_o values). It is likely that, for the gassy soils of Gardner,[14] the frequency

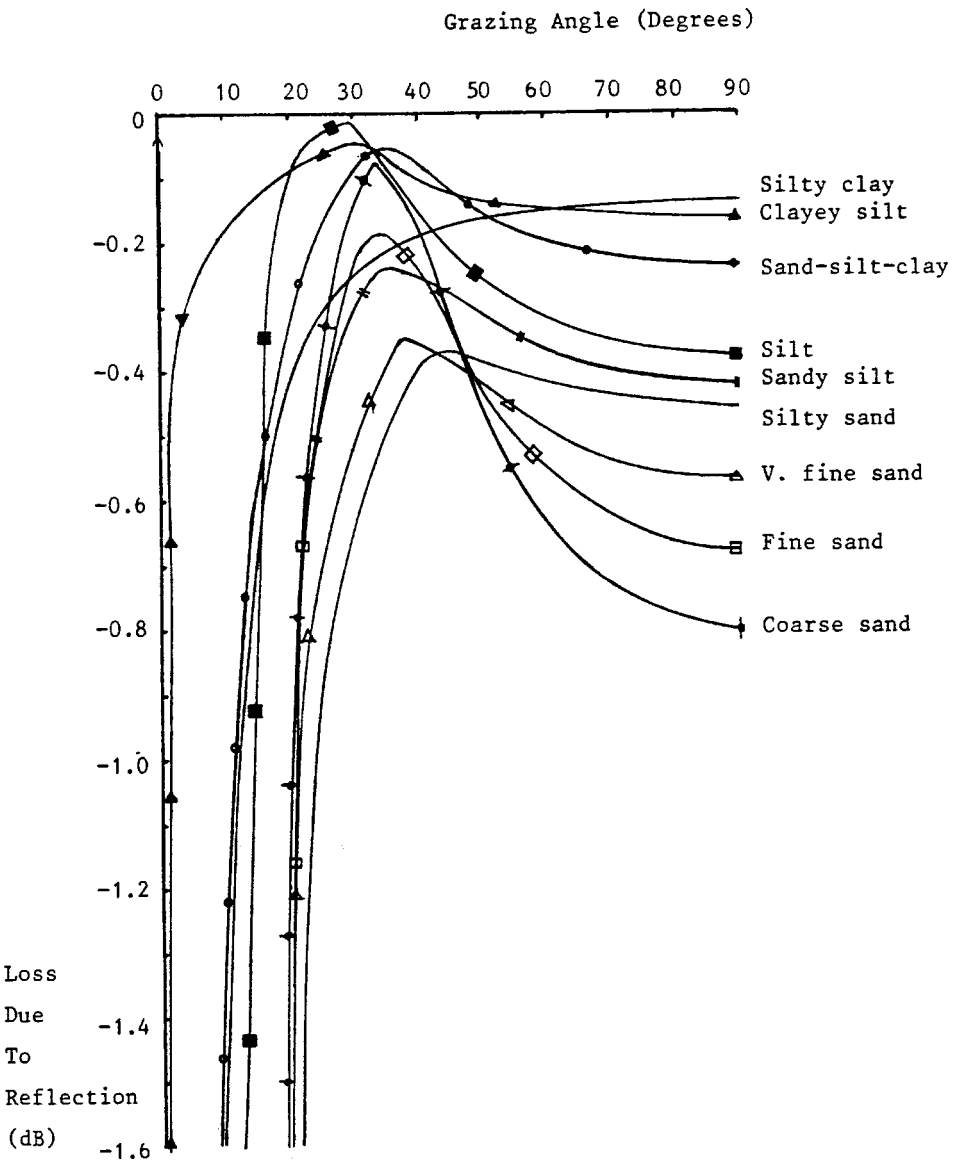


FIG. 6. Comparative reflection losses for nongassy sediments (silty clays to course sands). These are predicted by Mackenzie's model (Ref. 21), based on physical properties of saturated sediments reported by Hamilton (Refs. 25, 28). The absence of a critical angle, below which all energy is reflected, is typical of a silty clay for which the sound speed is generally less than in the overlying water. (Note that symbols do not denote data points; actual points are at degree intervals.)

ranges affected by bubble resonance (Fig. 2) are somewhat less than they are for water because of the additional damping effects caused by interparticle friction.

In summary, Anderson and Hampton's model for attenuation appears valid for the lower frequencies of bubble resonance and above frequencies of resonance; however, the model has not been shown to be reliable for frequencies below resonance. This may indicate that Anderson and Hampton's simulation of the bulk properties of the soil is in error, as was previously demonstrated for sound speed by Gardner and Sills.[13] They pointed out that, under acoustic loading, gassy soil responds as a two-phase medium constructed of gas pockets within a saturated soil matrix (of solid particles and fluid). It is suspected that the Anderson and Hampton model does not simulate this behavior, since it was developed from acoustic expressions derived from the behavior of a compressible fluid.

With respect to reflection losses, the sign reversal observed in Fig. 4 between E53-1 and E53-2 is the cause of the 180° phase reversal apparent in seismic records,[3] where profiling passes from a saturated sediment mudline to a gassy sediment. This of course arises from the acoustic impedance of gassy sediment being greater than fluid impedance, because, although the density of gassy sediment is greater than the overlying fluid density (Table I), sediment velocity is much lower than fluid velocity.[14] If we ignore interface scatter at the mudline which is not substantial,[8,32] the magnitudes of loss of up to 6 dB at 40° do not, however, substantiate speculation by Levin[7] of highly reflective pressure-release boundaries caused by low sediment velocities and high absorption losses in the gassy sediment interface. By contrast, the relatively high attenuation of the Gardner soils suggests that losses arise from absorption during signal transmission through the gassy sediments, and not by reflection at the mudline. Since absorption loss for frequencies below bubble resonance is several orders of magnitude higher than reflection losses, once seismic signals penetrate the mudline of gassy sediments they are much more adversely affected by transmission through the sediment. This is of concern for seismic surveys in gassy sediments similar to the Gardner soils and at frequencies below bubble resonance (10 kHz). The lightly gassy soils of gas fraction 0.003 57 (E51-2) will

respond largely as nongassy silty clay sediments with reflection losses of 0.15 dB (Fig. 6), and attenuations of less than 0.006 dB/cm.[33] These sediments may be penetrated easily to a significant depth. Conversely, the very gassy sediments of gas fraction 0.2 (E50-1) with reflection losses of 3 dB (Fig. 4) and attenuations of 20–25 dB/cm (Fig. 1) will be problematic. Although such mudlines may be penetrated easily, the substantial absorption loss arising from signal transmission will prevent penetration of the surficial layers to beyond 1 m in depth. Unfortunately, the more substantial losses for signal frequencies in gassy soils at bubble resonance, or in saturated soils above bubble resonance, suggest frequencies below resonance should be used for seismic profiling.

ACKNOWLEDGMENTS

The author would like to thank Professor Gilliane Sills for support and Mr. Christopher Waddup for construction of equipment, both from the Department of Engineering Science, University of Oxford.

APPENDIX A: PARAMETERS USED IN ANDERSON AND HAMPTON MODEL

$$A=[1+B^2]\left[1+\frac{3(\gamma-1)}{X}\left(\frac{\sinh X-\sin X}{\cosh X-\cos X}\right)\right],$$

$$B=3(\gamma-1)\times\left[\frac{X(\sinh X+\sin X)-2(\cosh X-\cos X)}{X^2(\cosh X-\cos X)+3(\gamma-1)(\sinh X-\sin X)}\right],$$

$$X=r_o\omega\rho g\frac{S_p}{c_g}.$$

ω=rotational frequency ($=2\pi f$)
f_o=resonance frequency
f=frequency
ρ_s=sediment matrix bulk density
S_p=specific heat of gas at constant pressure
c_g=conductivity of gas
r=bubble radius

$$f_o=\frac{1}{2\pi r}\sqrt{\frac{3\gamma P_o}{A\rho_s}+\frac{4G}{\rho_s}},$$

$$X_1=\sum_{l=1}^{M}\frac{v_l(1-f_{*l}^2)}{(1-f_{*l}^2)+d_{*l}^2},$$

$$Y_1=\sum_{l=1}^{M}\frac{v_l d_{*l}^2}{(1-f_{*l}^2)+d_{*l}^2},$$

where:

$f_{*1}=f/f_{o1}$
$d_{*1}=df_{*1}^2$
$d_r=kr[\omega_o/\omega]$
ω_o=rotational frequency at resonance ($=2\pi f_o$)
k=wave number ($=2\pi/\lambda$)
λ=wave length ($=c/f$)
$d_t=B$, the thermal damping constant
$d_f=4G'/(\rho_s\omega^2r^2)$, the friction/viscous damping constant
$d=d_r+d_t+d_f$, the damping constant
$G^*=G+iG'$, the complex shear modulus.
Sound speed (c):

$$\left(\frac{c_o}{c}\right)^2=\frac{1}{2}(1+WX_1)\left\{1\pm\left(1+\left[\frac{WY_1}{1+WX_1}\right]^2\right)^{1/2}\right\},$$

where c_o is the sound speed in saturated sediment, X_1 and Y_1 are coefficients of a gassy medium for which bubble-size distribution may be expressed as a histogram. Also, $W=K_{\text{sat}}/[\gamma P_o+(4G_{\text{sat}}/3)]$, γ is the specific heat ratio for gas (adiabatic constant), P_o is the ambient hydrostatic pressure, and K_{sat} and G_{sat} are, respectively, the dilatational and shear moduli of the soil matrix (*not* the *bulk* material properties).

1. Case i—Anderson and Hampton

Ratio of specific heats of gas in bubbles $\gamma=1.333$ (Kaye and Laby[34])

Reciprocal of thermal diffusivity of gas for bubbles in water (where for the gas s_p is specific heat at constant pressure, ρ_g is density, and C_g is thermal conductivity)—Anderson and Hampton[8,33]

$$\frac{\rho_g s_p}{C_g}=500\ \text{s/m}.$$

Soil matrix bulk density $\rho_s=1597\,\text{kg/m}^3$ (measured from saturated soil sample, Gardner[35])

Water density $\rho_w=1020\,\text{kg/m}^3$ (measured from water sample, Gardner[35])

Gas density $\rho_g=0.6623\,\text{kg/m}^3$ (CRC Handbook[36])

Sound speed in bubble-free water $c_o=1502.5$ (measured, Gardner[35])

Soil shear modulus $G_{\text{sat}}=1.3\times10^8\,\text{N/m}^2$ (Measurement from saturated surficial sediment—Smith[24] and Hamilton[25])

Imaginary part of complex modulus $G'=0.325\times10^8\,\text{N/m}^2$ (25% of G—Anderson and Hampton[8,33])

Soil dilatational modulus $K_{\text{sat}}=3.649\,86\times10^9\,\text{N/m}^2$ (from Gassman[26] equations—saturated medium)

Soil frame bulk modulus $K_f=0.007\,97\times10^{10}\,\text{N/m}^2$ saturated soil (Anderson and Hampton[8,33])

Mineral particle bulk modulus $K_m=5.0\times10^{10}\,\text{N/m}^2$ saturated soil (Anderson and Hampton[8,33])

Water bulk modulus $K_w=0.23\times10^{10}\,\text{N/m}^2$ saturated soil (Anderson and Hampton[8,33])

Ambient hydrostatic pressure $P_o=1.013\,25\times10^5\,\text{N/m}^2$

Total void porosity $n=0.625$ (measured from saturated soil sample, Gardner[35])

2. Case ii—Wheeler (Refs. 15–17) and Jardine *et al.* (Ref. 18)

As above, except for G_{sat} and K_{sat} as follows:

Soil shear modulus $G_{\text{sat}}=0.5\times10^8\,\text{N/m}^2$ ($G'_{\text{sat}}=0.125\times10^8\,\text{N/m}^2$)

Soil dilatational modulus $K_{\text{sat}}=4.211\times10^9\,\text{N/m}^2$.

APPENDIX B: THE MACKENSIE MODEL (Ref. 21) FOR REFLECTION LOSS

Let $s=(\rho_1 c_1 \sin\theta)/(\rho_2 c_2 \sin\phi)$; hence, the Rayleigh reflection coefficients are

$$\mathrm{R}=\frac{p_\mathrm{r}}{p_\mathrm{i}}=\frac{1-s}{1+s}$$
$$\mathrm{T}=\frac{p_\mathrm{t}}{p_\mathrm{i}}=\frac{2}{1+s}. \tag{B1}$$

Let $a_1=k_1\sin\phi$, $b_1=k_1\cos\phi$, $a_2=k_2\sin\theta$, and $b_2=k_2\cos\theta$, where $k_1=w/c_1$ and $k_2=w/c_2$, $w=2\pi f$

$$s=\frac{\rho_1 a_2}{\rho_2 a_1}.$$

To account for absorption in the sediment surface, let $k_2=\beta+\mathrm{i}\alpha$, where α and β are real and positive.

For the boundary conditions $b_2=b_1$ and a_2 and c_2 are complex,

$$\text{impedance ratio } \sigma=\frac{\rho_2 k_1}{\rho_1\beta}=\frac{\rho_2 v_2}{\rho_1 c_1},$$

where phase velocity $v_2=w/\beta$.

Index of refraction $n=\beta/k_1=c_1/v_2$ (for a nondispersive medium, $v_2=c_2$).

Let $a_2/\beta=h+\mathrm{i}g$, then $s=(h+\mathrm{i}g)/\sigma\sin\phi$.

The squares of R and T are therefore Eq. (B1) multiplied by their complex conjugates

$$\left|\frac{p_\mathrm{r}}{p_\mathrm{i}}\right|^2=\frac{(1-s)(1-s^*)}{(1+s)(1+s^*)}=\frac{(h-\sigma\sin\phi)^2+g^2}{(h+\sigma\sin\phi)^2+g^2}$$
$$\left|\frac{p_\mathrm{t}}{p_\mathrm{i}}\right|^2=\frac{4}{(1+s)(1+s^*)}=\frac{4\sigma^2\sin^2\phi}{(h+\sigma\sin\phi)^2+g^2}. \tag{B2}$$

Hence, loss due to reflection [using Eq. (2), Sec. II B] is

$$10\log_{10}\left[\left|\frac{p_\mathrm{t}}{p_\mathrm{i}}\right|^2\frac{\rho_1 c_1}{\rho_2 c_2}\right], \tag{B3}$$

where

$$\beta=k_1\frac{c_1}{c_2}=\frac{w}{c_2}=\frac{2\pi f}{c_2},$$

$$\sigma=\frac{\rho_2 c_2}{\rho_1 c_1},$$

$$\mathrm{A}=\frac{\alpha}{\beta},$$

$$\mathrm{B}=\frac{1}{2}\left[1-\left(\frac{\cos\phi}{n}\right)^2-\left(\frac{\alpha}{\beta}\right)^2\right]$$

$$n=\frac{c_1}{c_2}$$

$$h=(\mathrm{B}+(\mathrm{A}^2+\mathrm{B}^2)^{1/2})^{1/2}$$

$$g=(-\mathrm{B}+(\mathrm{A}^2+\mathrm{B}^2)^{1/2})^{1/2}$$

Notation

p_r, p_t, p_i are pressure amplitudes of reflected, transmitted, and incident waves respectively
f=frequency
σ=impedance ratio
c_1, c_2, ρ_1, ρ_2 are sound speeds and bulk densities of the water and sediments, respectively
ϕ, θ are grazing angles of incident and refracted waves, respectively
s^*=complex conjugate of s
α=attenuation in the sediment

[1] A. L. Anderson, F. Abegg, J. A. Hawkins, M. E. Duncan, and A. P. Lyons, "Bubble populations and acoustic interaction with the gassy floor of Eckernforde Bay," Cont. Shelf Res. **18**, 1807–1838 (1998).
[2] H. E. Egerton, "Sub-bottom penetration in Boston Harbor," J. Geophys. Res. **70**, 2931–2934 (1965).
[3] N. G. T. Fannin, "I.G.S. pockmark investigations 1974–1978," Institute of Geological Sciences, IGS Report No. **98** (1979).
[4] T. N. Gardner and M. J. Goringe, "The measurement of gas bubble size distribution in a three phase laboratory gassy soil," Geotech. Test. J. ASTM **11**, 49–55 (1988).
[5] J. R. Schubel, "Gas bubbles and the acoustically inpenetrable, or turbid, character of some estuarine sediments," edited by I. R. Kaplan (Plenum, New York, 1974).
[6] G. A. Ruff, "Acoustic characteristics of Black Moshannon Lake bottom," J. Acoust. Soc. Am. **42**, 524–525 (1967).
[7] F. K. Levin, "The seismic properties of Lake Maracaibo," Geophysics **27**, 35–47 (1962).
[8] A. L. Anderson and L. D. Hampton, "Acoustics of gas bearing sediments. II. Measurements and models," J. Acoust. Soc. Am. **67**, 1890–1903 (1980).
[9] A. Bedford and M. Stern, "A model for wave propagation in gassy sediments," J. Acoust. Soc. Am. **73**, 409–417 (1983).
[10] T. S. Edrington and T. M. Calloway, "Sound speed and attenuation measurements in gassy sediments in the Gulf of Mexico," Geophysics **49**, 297–299 (1984).
[11] W. L. Nyborg, I. Rudnick, and H. K. Schilling, "Experiments on acoustic absorption in sand and soil," J. Acoust. Soc. Am. **22**, 422–425 (1950).
[12] A. P. Lyons, M. E. Duncan, A. L. Anderson, and J. A. Hawkins, "Predictions of the acoustic scattering response of free-methane bubbles in muddy sediments," J. Acoust. Soc. Am. **99**, 163–172 (1996).
[13] T. N. Gardner and G. C. Sills, "An examination of the parameters that govern the acoustic behavior of sea bed sediments containing gas bubbles," J. Acoust. Soc. Am. **110**, 1878–1889 (2001).
[14] T. N. Gardner, "An acoustic study of soils that model sea-bed sediments containing gas bubbles," J. Acoust. Soc. Am. **107**, 163–176 (2000).
[15] S. J. Wheeler, "A conceptual model for soils containing large gas bubbles," Geotechnique **38**(3), 389–397 (1988).
[16] S. J. Wheeler, "The undrained shear strength of soils containing large gas bubbles," Geotechnique **38(3)**, 399–413 (1988).
[17] S. J. Wheeler, "The stress strain behaviour of soils containing gas bubbles," Ph.D. thesis, University of Oxford (1986).
[18] R. J. Jardine, M. J. Symes, and J. B. Burland, "The measurement of soil stiffness in the triaxial apparatus," Geotechnique **34(3)**, 323–340 (1984).
[19] E. Silberman, "Sound velocity and attenuation in bubbly mixtures measured in standing wave tubes," J. Acoust. Soc. Am. **29**, 925–933 (1957).
[20] J. D. Macpherson, "Effect of gas bubbles on sound propagation in water," Proc. Phys. Soc. London, Sect. B **70**, 85–92 (1957).
[21] K. V. Mackenzie, "Reflection of sound from coastal bottoms," J. Acoust. Soc. Am. **32**, 221–231 (1960).
[22] J. W. Rayleigh, *The Theory of Sound* (Strutt, 1877).
[23] T. G. Muir, C. W. Horton, and L. A. Thomson, "The penetration of highly directional acoustic beams into sediments," J. Sound Vib. **64(4)**, 539–551 (1979).
[24] D. T. Smith, "Acoustic and mechanical loading of marine sediments," ONR Symposium on "Physics of sound in marine sediments," Lakeway Inn, Austin, TX, (1973).

[25] E. L. Hamilton, ''Elastic properties of marine sediments,'' J. Geophys. Res. **76**, 579–604 (1971).
[26] F. Gassman, ''Elastic waves through a packing of spheres,'' Geophysics **16**, 673–685 (1951).
[27] S. J. Wheeler and T. N. Gardner, ''Elastic moduli of soils containing large gas bubbles,'' Geotechnique **39**(2), 333–342 (1989).
[28] E. L. Hamilton and R. T. Bachman, ''Sound velocity and related properties of marine sediments,'' J. Acoust. Soc. Am. **72**, 1891–1904 (1982).
[29] A. B. Wood and D. E. Weston, ''The propagation of sound in mud,'' Acustica **14**, 156–162 (1964).
[30] A. L. Anderson, R. J. Harwood, and R. T. Lovelace, ''Investigation of gas in bottom sediments,'' Applied Research Laboratories, University of Texas, Austin, Report ARL-TR-70-28 (1971).
[31] E. L. Hamilton, ''Compressional-wave attenuation in marine sediments,'' Geophysics **37**(4), 620–646 (1972).
[32] A. L. Anderson and L. D. Hampton, ''Acoustics of gas bearing sediments. I. Background,'' J. Acoust. Soc. Am. **67**, 1865–1889 (1980).
[33] E. L. Hamilton, ''Sound channels in surficial marine sediments,'' J. Acoust. Soc. Am. **48**, 1269 (1970).
[34] J. V. C. Kaye and T. H. Laby, *Tables of Physical and Chemical Constants*, 14th ed. (Longmans, New York, 1973).
[35] T. N. Gardner, ''The acoustic properties of gassy soil,'' Ph.D. engineering thesis, University of Oxford (1988).
[36] *CRC Handbook of Chemistry and Physics* (CRC Press, Boca Raton, FL, 1999).

Improvement in matched field processing using the CLEAN algorithm

H. C. Song, J. de Rosny,[a] and W. A. Kuperman
Marine Physical Laboratory, Scripps Institute of Oceanography, University of California–San Diego, La Jolla, California 92093-0238

(Received 13 August 2001; revised 8 February 2002; accepted 17 June 2002)

Adaptive matched field processing such as the minimum variance distortionless processor (MV) provides excellent sidelobe (or ambiguity) suppression capability in source localization given a perfect knowledge of the ocean environment. Unfortunately, this processing is very sensitive to sources of mismatch and robust adaptive algorithms are then employed such as a white noise constraint (WNC) often at the expense of insufficient sidelobe control. The CLEAN algorithm was introduced in radio astronomy [Astron. Astrophys. Suppl. Ser. **15**, 417–426 (1974)] to produce a high quality image of the sky by reducing sidelobe-induced artifacts. In this paper, the CLEAN concept is extended to matched field processing. Numerical simulations and experimental data demonstrate that matched field processing combined with the CLEAN algorithm can improve performance, especially when a weak source is masked by sidelobes from a much stronger source.
 [DOI: 10.1121/1.1531510]

PACS numbers: 43.30.Wi, 43.60.Gk [PLB]

I. INTRODUCTION

In this paper we demonstrate that matched field processing (MFP) in the ocean[1] can be improved by postprocessing with the CLEAN algorithm,[2] most widely used in radio astronomy to produce a high quality image of the sky. The CLEAN algorithm, introduced by Högbom for synthesis imaging, assumes that the radio sky can be represented by a small number of point sources in an otherwise empty field of view. Then it uses a simple iterative procedure to find the positions and strengths of these sources. On convolving the point-source components with a so-called ''CLEAN'' beam, a final image called a ''CLEAN'' image is obtained. The CLEAN algorithm was designed to reduce the sidelobe-induced artifacts due to incomplete sampling of the radiation fields derived from independently radiated point sources such as stars.[3]

Matched field processing in the ocean used to localize discrete point sources in range, depth and azimuth where the output of the processor is often refered to as an ambiguity surface or an image. MFP, a generalization of classical plane wave beam forming, exploits the spatial complexity of the ocean environment.[1] When there are multiple discrete point sources in the image, a nonlinear adaptive processor such as the minimum variance distortionless processor (MV) provides excellent sidelobe suppression capability as shown in Fig. 2. Figure 1 is a schematic for the MFP example with two targets in the presence of surface generated noise.[5] Given a perfect knowledge of the environment, the MV (top middle) clearly identifies the two targets in red circle with high resolution, whereas the conventional Bartlett processor (top left) localizes only the stronger one. Once a slight mismatch is introduced (e.g., sound speed in Fig. 3), the MV localizes neither of the two targets (bottom middle) due to its sensitivity to mismatch. In this case, the more robust white noise constraint (WNC) adaptive processor[4] (bottom right) allows a localization of the stronger target along with a few sidelobes in the ambiguity surface including a weak target. The objective of this paper is to extend the concept of the CLEAN algorithm to MFP to see if we can improve its localization capability further by postprocessing the WNC ambiguity surface (bottom right).

Section II discusses the basic concept and procedures of the CLEAN algorithm in radio astronomy. Section III describes how the CLEAN algorithm can be adopted in MFP. Section IV shows examples of numerical simulations as well as an example applied to real experimental data collected during SWellEx-96, followed by summary.

II. THE CLEAN ALGORITHM

Details of CLEAN algorithm in synthesis imaging can be found in Refs. 6 and 7. In this section, we review briefly the basic concept of synthesis imaging to describe the CLEAN algorithm.

Synthesis imaging, using the continuous rotation of the Earth, creates images of astronomical sources from arrays of radio telescopes. The Van Cittert–Zernicke theorem[8] relates the observed visibility function $V(u,v)$ and source brightness (or intensity) $I(l,m)$,

$$I(l,m)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}V(u,v)e^{2\pi i(lu+mv)}\,du\,dv=F[V(u,v)], \tag{1}$$

where F indicates the Fourier transform, (u,v) is a separation in wavelengths in two-dimensional space, and (l,m) is an angle expressed as a direction cosine. The visibility function $V(u,v)$ is also referred to as the spatial coherence function of the electric field E, as measured by radio interferometers. In practice the visibility function $V(u,v)$ is not known ev-

[a]Current address: Laboratoire Ondes et Acoustique, E.S.P.C.I., Université Paris VII, 10 rue Vauquelin, 75005 Paris, France.

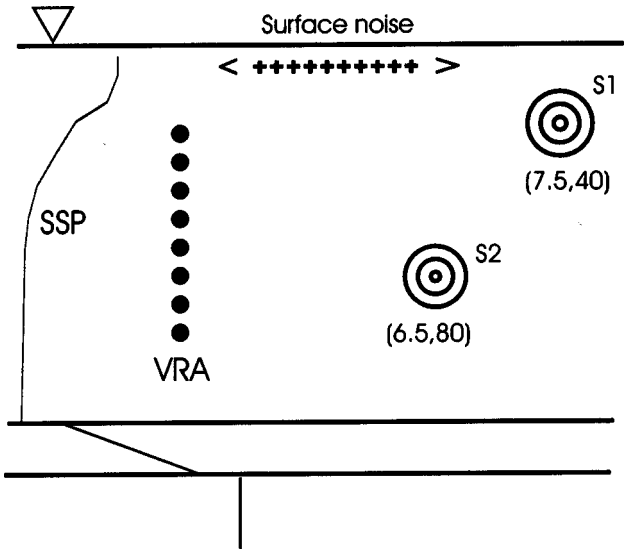


FIG. 1. Schematic of the MFP example. A vertical array consists of 29 elements with the uppermost element at 15 m depth and the element spacing of 2.786 m. The sound speed displays a downward refracting profile in the water depth of 114 m. There are two discrete sources: a stronger source with 110 dB SL (source level) at 7.5 km range and 40 m depth from the VRA, a weaker source with 100 dB SL at 6.5 km range and 80 m depth. The surface generated noise with 60 dB strength (Ref. 5) is added to the cross spectral density matrix (CSDM). The source frequency is 450 Hz.

erywhere but is sampled only at discrete points described by a sampling function $S(u,v)$. So only an approximation to I, called a ''dirty map'' and denoted by I', is obtained. Using the convolution theorem for Fourier transforms, it is then given by

$$I'(l,m)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}V(u,v)S(u,v)e^{2\pi i(lu+mv)}\,du\,dv$$
$$=F[VS]=I*F[S]=I*B, \tag{2}$$

where * denotes a convolution. The Fourier transform of the sampling function S is B defined as the ''dirty beam'' or synthesized beam (or point spread function). A dirty map I' is also called the principal solution of the image with all unsampled visibilities set to zero. The quality of the resulting image depends entirely on sampling in the (u,v) plane.

Because linear deconvolution algorithms such as Wiener filtering and inverse filtering are not applicable to observations in synthesis imaging due to nonuniform discrete sampling, nonlinear algorithms such as CLEAN or MEM (maximum entropy method) have been developed in radio astronomy.[7] Note that the solution to the convolution equation Eq. (2) is not unique and the problem of image reconstruction is reduced to that of choosing a plausible image from the set of possible solutions. We confine our interest in the most widely used CLEAN algorithm and the MEM approach is discussed in Ref. 9. The CLEAN algorithm, first introduced by Högbom,[2] is an iterative method which deconvolves a sampling function B (the dirty beam) from an observed brightness I' (dirty map) of a radio source. There are many variants of the basic CLEAN algorithm available which extend the method to achieve greater speed and produce more realistic maps.[7] Here we describe the original Högbom CLEAN algorithm along with its terminology: Dirty Map (initial image), Dirty Beam, CLEAN window, CLEAN component, residual, loop gain, CLEAN Beam (restoring beam) and CLEAN Map (final image). Note that a map or an image is interchangeably used throughout the paper.

The iterative procedure operates in the image plane and begins with a dirty map $I'(l,m)$ from which a set of point sources $P(l,m)$ (CLEAN components) is derived, such that when convolved with the dirty beam $B(l,m)$ the resulting image is almost equal to the dirty map except for a residual $\Delta(l,m)$ defined by

$$\Delta(l,m)=I'(l,m)-\int\int P(l',m')$$
$$\times B(l-l',m-m')dl'\,dm'. \tag{3}$$

The procedure consists of the following steps:[6]

(1) Locate the maximum in the dirty map $I'(l,m)$ and determine its amplitude P.
(2) Convolve the dirty beam $B(l,m)$ with a point source (at this location) of amplitude γP, where γ is known as the loop gain ($\gamma\leqslant 1$).
(3) Subtract the result of this convolution from the dirty map.
(4) Repeat until the residual $\Delta(l,m)$ is sufficiently small or an iteration limit N is reached.
(5) Convolve the point source collection with a CLEAN beam such as a Gaussian function to obtain a smooth, clean map.
(6) Add the residual to the CLEAN map.

The last step is sometimes omitted, but is recommended because it can provide useful diagnostic information about the noise on the image, residual sidelobes, ''bowls'' near the image center. In step 3, a positivity constraint can be applied if the map is to be the total brightness of a source.

Steps 1, 2, 4, and 5 are at the control of the map maker. If we happen to know that the image lies within certain *a priori* known regions or boxes in the field, then we may limit the search to these boxes, called a CLEAN window. If the image has compact components only, the loop gain γ can be as large as unity so as to reduce the number of iterations. If the image is extended and smooth, a small loop gain will be necessary and more iterations are required.[6] Finally, the choice of the CLEAN beam will affect the resulting clean image. The most common way to choose the CLEAN beam is to fit an elliptical Gaussian to the main lobe of the dirty beam, a compromise between resolution and apparent image quality.[7,10] It is appropriate to mention that although CLEAN is extensively used in radio astronomy, theoretical understanding is relatively poor[11,12] and the existence of instability is well-known[13] in addition to its nonuniqueness. The condition for the convergence of CLEAN is discussed in Ref. 11.

III. MFP WITH THE CLEAN ALGORITHM

An overview on matched field processing (MFP) is given by Baggeroer *et al.* in Ref. 1. We briefly review the relevant processors that will be employed in the CLEAN

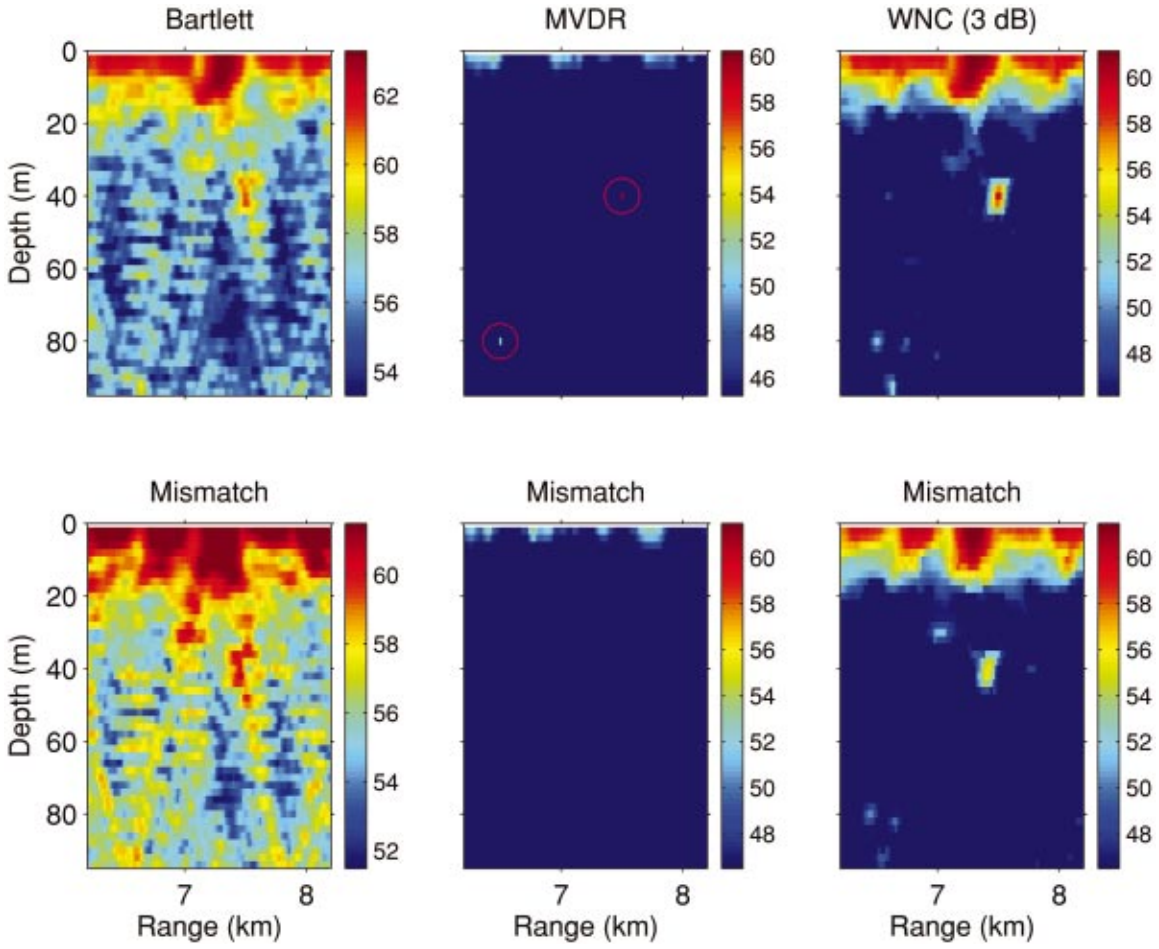


FIG. 2. Matched field processing with a perfect knowledge of the environment (top row) and mismatch in sound speed (bottom row): Bartlett (left), MV (middle), and WNC (right). Given a perfect knowledge of the environment, the adaptive MV processor (top middle) clearly identifies two targets with excellent sidelobe control and there is no need for a robust adaptive processor (top right). With a slight mismatch in sound speed (see Fig. 3), however, MV (bottom middle) can localize neither of the two targets. On the other hand, WNC (bottom right) processor with 3 dB constraint localizes the stronger target at 7.5 km range and 40 m depth along with several sidelobes. The WNC ambiguity surface can be used as a primary or ''dirty'' image for the CLEAN process later. The dynamic range of the ambiguity surface is 15 dB.

process, describe the CLEAN procedures as applied to MFP, and compare MFP in ocean acoustics with systhesis imaging in radio astronomy.

Matched field processing is a generalized beamforming method which uses the spatial complexities of acoustic fields in an ocean waveguide to localize sources in range, depth and azimuth or to infer parameters of the waveguide itself. Matched field processing matches the measured field at the array with replicas of the expected field for all possible source locations. In MFP, the measured data is the cross spectral density matrix (CSDM) in the frequency bin of interest from the time series at a receiving vertical array in a wave guide, which include the signal and noise fields. Assuming the signal and noise are independent statistically, the total CSDM is the outer product of data vectors, say $\mathbf{d}$

$$\mathbf{K}=\mathbf{E}[\mathbf{d}\mathbf{d}^{\dagger}]=\mathbf{K}_s+\mathbf{K}_n, \tag{4}$$

where $\mathbf{K}_s$ and $\mathbf{K}_n$ are the CSDM of the signal and the noise, respectively. In practice the estimation of $\mathbf{K}$ is done by averaging over a number of snapshots. The noise CSDM $\mathbf{K}_n$ consists of white noise, distributed noise (e.g., surface-generated noise), and discrete noise sources.

Conventional (Bartlett) linear MFP is performed by weighting the output of the array elements and summing over all elements. The expected value of the power output of the beamformer is written as the quadratic form († denotes the complex transpose operation)

$$P_{BT}(\hat{\mathbf{a}})=\mathbf{w}^{\dagger}(\hat{\mathbf{a}})\mathbf{K}(\mathbf{a}_T)\mathbf{w}(\hat{\mathbf{a}}), \tag{5}$$

where $\hat{\mathbf{a}}$ is the scanning parameters of the source location for which we construct the replica fields $\mathbf{w}(\mathbf{a})$. The parameter vector $\mathbf{a}$ incorporates all source coordinates and $\mathbf{a}_T$ is the true parameter of the source location. The weighting vector of the array element output $\mathbf{w}(\hat{\mathbf{a}})$ depends upon the replica field and possibly the noise environment. For conventional non-adaptive processing, the weight or projection vectors, all have unit length, i.e.,

$$\mathbf{w}(\mathbf{a})=\Phi(\mathbf{a})/|\Phi(\mathbf{a})|, \tag{6}$$

where $\Phi(\mathbf{a})$ is a vector whose element is the solution to the wave equation for a source located at $\mathbf{a}$.

As compared to conventional Bartlett processors, the minimum variance distortionless processor (MV) is a high

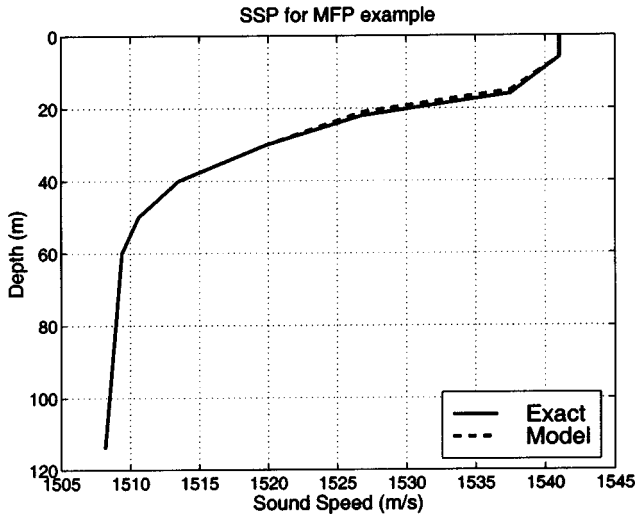


FIG. 3. Sound speed profile for numerical simulation in Fig. 4: exact (solid line) and mismatched profile (dotted line). The mismatch in sound speed occurs around 15–30 m depth due to a slight difference in gradient.

resolution array processing algorithm that adaptively constructs a weighting (or steering) vector such that

$$\mathbf{w}_{\mathrm{MV}}(\hat{\mathbf{a}}) = \mathbf{K}^{-1}\mathbf{w}(\hat{\mathbf{a}}) / \mathbf{w}^{\dagger}(\hat{\mathbf{a}})\mathbf{K}^{-1}\mathbf{w}(\hat{\mathbf{a}}), \tag{7}$$

where $\mathbf{w}(\hat{\mathbf{a}})$ is derived as in Eq. (6). Then the MV power output is obtained by replacing $\mathbf{w}$ by $\mathbf{w}_{\mathrm{MV}}$ in the quadratic form of Eq. (5) and carrying out the matrix algebra,

$$P_{\mathrm{MV}} = |\mathbf{w}^{\dagger}(\hat{\mathbf{a}})\mathbf{K}^{-1}\mathbf{w}(\hat{\mathbf{a}})|. \tag{8}$$

MVDR is an optimum processor explained in terms of the ''estimator-subtractor'' attribute that adaptively attempts to cancel sidelobes assuming a perfect knowledge of the environment is given.[14]

It is well-known, however, that MV is very sensitive to mismatch.[15] Thus a more robust processor has been developed, which incorporates a white noise constraint (WNC), a quadratic inequality constraint.[4] In this case the steering vector $\mathbf{w}$ takes the form

$$\mathbf{w}_{\mathrm{WNC}}(\hat{\mathbf{a}}) = (\mathbf{K} + \varepsilon\mathbf{I})^{-1}\mathbf{w}(\hat{\mathbf{a}}) / \mathbf{w}^{\dagger}(\hat{\mathbf{a}})(\mathbf{K} + \varepsilon\mathbf{I})^{-1}\mathbf{w}(\hat{\mathbf{a}}), \tag{9}$$

where a small amount is added to the diagonal elements of the CSDM (i.e., $\varepsilon\mathbf{I}$). Note that as ε approaches infinity, the CSDM becomes diagonally dominant, and the resulting weights are the same as those for the conventional processor (BT). On the other hand, WNC reduces to the MVDR in Eq. (7) as ε becomes zero. Therefore, WNC allows a trade-off between the robustness of the processor and resolution by an appropriate value of ε. As described in Fig. 2, however, the trade-off often cannot provide enough sidelobe suppression capability to localize a weaker target. The CLEAN algorithm then can be applied to the resulting ambiguity surface as a dirty map for potential improvement.

A. CLEAN process

The CLEAN algorithm can reduce sidelobes and increase target dynamic range. The CLEAN process is applied to MFP as follows:

(1) Generate a best ambiguity surface I' as possible using MFP which corresponds to a dirty or a primary image. In particular, we consider the robust WNC processor assuming there is a mismatch in the environment.
(2) Find the brightest (strongest) spot in the ambiguity surface, $\hat{\mathbf{a}}_{pk}$ (CLEAN component). The search for peaks may be constrained to a CLEAN window with *a priori* knowledge. For example, we can exclude near-surface region assuming that discrete point sources are in the deeper water column.
(3) Calculate a dirty beam B using a Bartlett processor with the CSDM of $\mathbf{K} = \Phi(\hat{\mathbf{a}}_{pk})\Phi^{\dagger}(\hat{\mathbf{a}}_{pk}) / |\Phi(\hat{\mathbf{a}}_{pk})|^2$, where $\Phi(\hat{\mathbf{a}}_{pk})$ is a replica vector. From the standpoint of robustness and resolution, it is desirable to average out the $\mathbf{K}$ matrix in the neighborhood of $\hat{\mathbf{a}}_{pk}$.
(4) Subtract from the dirty image, at the position of the peak, the dirty beam B multiplied by the peak strength and a loop gain $\gamma \leqslant 1$.
(5) Go to step 2 unless any remaining peak is below some user-specified level or a specified iteration limit N is reached.
(6) Generate CLEAN beams corresponding to the CLEAN components using any kind of MFP processor excluding sidelobes. For example, the mainlobe of the dirty beam B in step 3 can be used as a CLEAN beam.
(7) Add the residual of the dirty image to the CLEAN beams with its strength to create a final CLEAN image.

Note that there is one to one correspondence between the steps 2–7 above and the steps 1–6 in Sec. II. Although the dirty beams are generated differently in MFP in step 3. It is assumed that the brightest spot in the initial ambiguity surface is the response to a real target and not the accidental constructive interference of the sidelobe responses of several smaller targets. In step 4, a positivity constraint of the ambiguity surface should be applied such that all the negative points in the residual are replaced by zeros.

It should be mentioned that the postprocessing with the CLEAN algorithm does not require much computation load because dirty beams are generated by the fast Bartlett MFP and it involves only subtractions in the ambiguity surface.

B. Comparison between MFP and synthesis imaging

In this section we investigate the similarity and difference between MFP in ocean acoustics and synthesis imaging in radio astronomy. The fundamental assumption in both cases is that the image can be represented by a number of discrete point sources or stars although the CLEAN algorithm has been found to work well for extended sources as well in radio astronomy.[7] The measured raw data is the acoustic pressure field $p(r,z)$ at a vertical array in MFP, whereas the electromagnetic field $E(\mathbf{r})$ is measured by a radio telescope. The actual processed data for MFP is a cross spectral density matrix (CSDM) $\mathbf{K}$ while radio interferometer produces the spatial coherence function or visibility function V. The difference between the two is the mapping from the processed data to the image. As described in Sec. II, the source brightness in synthesis imaging $I(l,m)$ is the Fourier transform of the measured visibility function $V(u,v)$,

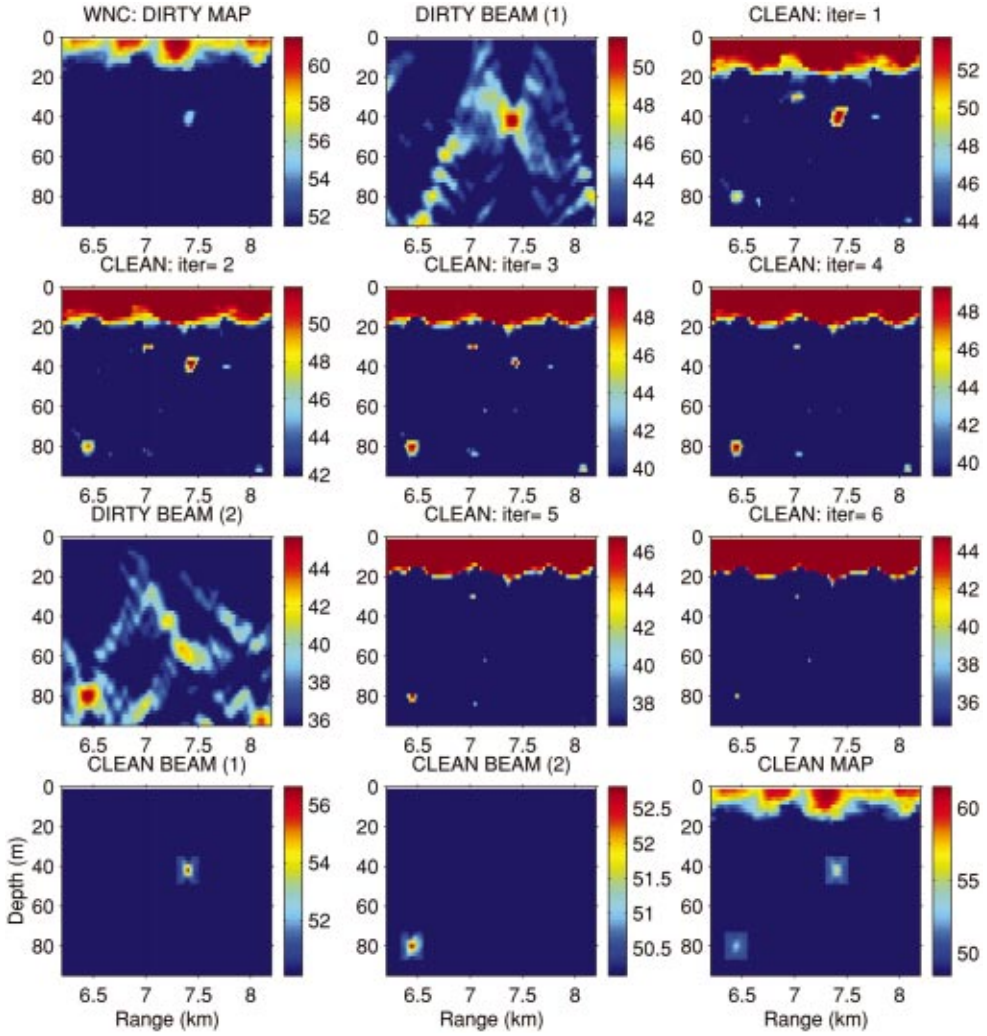


FIG. 4. A dirty map, two dirty beams, residuals, two CLEAN beams, and a CLEAN image. The initial dirty map is the ambiguity surface generated by a 3 dB white noise constraint (WNC) MFP processor equivalent to Fig. 2 (bottom right) with a bright spot at 7.5 km range and 40 m depth (first CLEAN component). Note that the dynamic range is decreased to 10 dB. A slight mismatch in sound speed is introduced as shown in Fig. 3. A dirty beam constructed from the Bartlett processor is shown next to the dirty map (top center). The loop gain γ is 0.6. After the fourth iteration, the residual finds the peak at the position of the weaker target at 6.4 km range and 80 m depth (second CLEAN component) with 100 m bias in range (true target at 6.5 km range). The second dirty beam is generated from this new peak and the cleaning process is continued until the remaining peak is below some noise level after the sixth iteration. The following two images are the CLEAN beams corresponding to the two CLEAN components (peaks), respectively. The mainlobe of the ambiguity surface from MV processor is chosen as the CLEAN beam. Finally, a CLEAN image is displayed in the bottom right by adding the two CLEAN beams with appropriate strengths to the residual of the sixth iteration. Note that the dynamic range is increased to 13 dB in order to show both targets and surface generated noise.

i.e., linear transform mapping from the van Cittert–Zernicke theorem. The major problem in image reconstruction is that of incomplete data and various methods have been developed including the CLEAN algorithm.[6] In MFP, however, the ambiguity surface, power output of the beam former P, includes not only measurement of CSDM $\mathbf{K}$ at the array, but also the weight or steering vector $\mathbf{w}(\hat{\mathbf{a}})$ based on *a priori* knowledge of the environment. The major problem in MFP arises from the incomplete knowledge of the environment and more robust algorithm have been developed.[1] In addition, the steering vector $\mathbf{w}(\hat{\mathbf{a}})$ is computed for each pixel with an optimum diagonal loading ε in WNC. Thus, the mapping in MFP is nonlinear. When the CLEAN algorithm is applied to MFP, however, it uses the WNC output as an initial dirty map and actually employs the linear Bartlett processor to generate the dirty beams, noting that the inner product with replica in Bartlett MFP corresponds to the Fourier transform in synthesis imaging.

IV. EXAMPLES

This section provides two examples of adaptive WNC MFP combined with the CLEAN process described in Sec. IV A. The first CLEAN example is a numerical simulation applied to the situation introduced earlier in Fig. 2. The second example uses experimental data recorded off San Diego in 1996.[16,17]

A. Numerical simulation

Figure 4 shows a complete picture of CLEAN procedure applied to MFP. It is assumed that the ensemble covariance

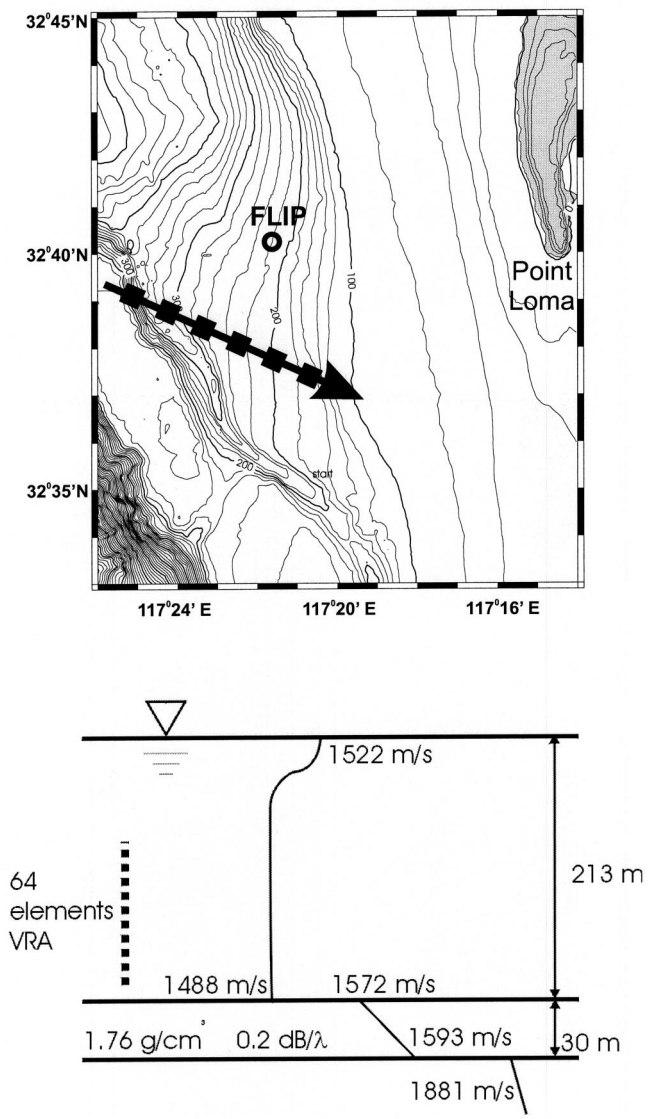


FIG. 5. SWellEx-96 Experimental area (top) off Point Loma, San Diego. A large ship traveled the area SW to SE of the R/P FLIP with a CPA approximately 3.5 km which occurred around 1215 Z, JD 131. The environmental model to generate replica fields for MFP is shown below.

K is known for this example. The initial dirty map (top left) is the ambiguity surface resulting from a 3 dB WNC processor as shown in Fig. 2 (bottom right). The range and depth grids Δr, Δz are 20 m and 2 m, respectively. A slight mismatch in sound speed is introduced around 15–30 m depth shown in Fig. 3. Note that the dynamic range is decreased to 10 dB here. The brightest spot in the image corresponds to the stronger target at 7.5 km range and 40 m depth (first CLEAN component). We constrain the CLEAN window or search space below 20-m depth since we are not interested in distributed surface noise. A dirty beam constructed from a Bartlett processor (top center) is shown next to the dirty map. The loop gain γ is chosen 0.6 for this example. After the 4th iteration, the residual finds the peak at the position of the weaker target at 6.4 km range and 80 m depth (second CLEAN component) with 100 m bias in range (true target at 6.5 km range). The second dirty beam is generated from this new peak and the cleaning process is continued until the remaining peak is below the ambient noise level after the sixth iteration. The following two images are the CLEAN beams corresponding to the two CLEAN components (peaks), respectively. The mainlobe of the ambiguity surface from MV processor is chosen as the CLEAN beam. Finally, a CLEAN image is obtained in the right bottom by adding the two CLEAN beams with appropriate strengths to the sixth residual. Note that the dynamic range is increased to 13 dB in order to show both targets and surface generated noise.

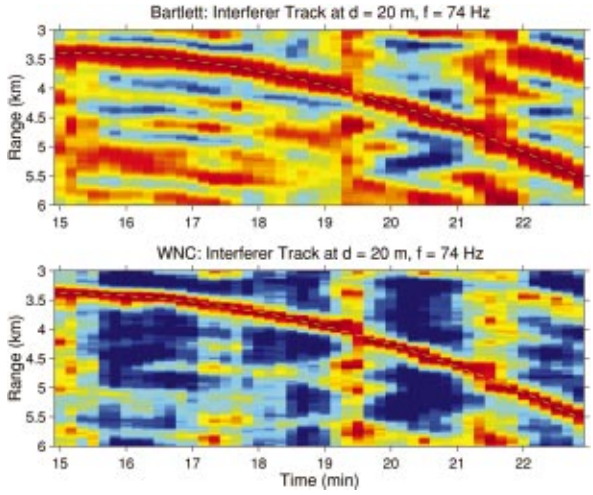


FIG. 6. A large ship track during SWellEx-96 Experiment off San Diego. The ship trajectory is superimposed with a dashed line (running from 3.4 km range at 15 minutes to 5.5 km at 23 minutes). The source depth is fixed at 20 m depth. The position of the ship is computed every 10 seconds using CSDM generated by 60 snapshots from 20 seconds of data with each snapshot FFT length of $N=1024$, Hanning window, and 50% overlap. WNC shows better performance with its sidelobe control.

B. Experimental data

1. SWellEx-96 experiment

The SWellEx-96 was carried out in May 1996 west of Point Loma in approximately 200 m deep water.[16,17] A vertical array data is analyzed when a large surface ship traveled the area SW to SE of the R/P FLIP (center) with a closest point of approach (CPA) of approximately 3.5 km around 1215 Z, JD131. Figure 5(a) shows a radar image of the ship track with square marks every 5 minutes and Fig. 5(b) is the environmental model to generate replica field for matched field processing.

Figure 6 shows the time-evolving ship track in range by a Bartlett processor (top) and WNC processor with 3 dB constraint (bottom). The source frequency of 74 Hz is chosen due to higher signal to noise ratio. The source depth is fixed at 20 m and the ship trajectory is a dashed line running from 3.4 km range at 15 minutes to 5.5 km at 23 minutes. The source range is updated every 10 seconds (frame) and each CSDM is estimated from 60 snapshots out of 20 seconds of data with each snapshot of FFT length $N=1024$, Hanning windowed with 50% overlap between successive FFTs. As expected, the WNC shows better performance with its sidelobe control. The sampling frequency was 1500 Hz.

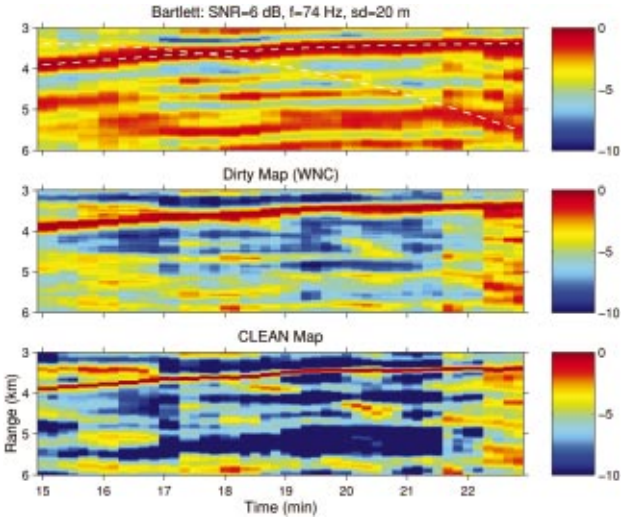


FIG. 7. Ambiguity surface with two-target scenario: Bartlett (top), WNC middle), and CLEAN map (bottom). The intensity of the second target moving from 4 km to 3.3 km with a half speed is 6 dB higher than the original ship track (from 3.4 km to 5.5 km). Two ship trajectories are superimposed as long-dashed lines in the top panel. As expected, WNC processor (middle) shows better performance than the conventional Bartlett processor (top), but it still does not show the weaker ship track. The cleaning process applied to the WNC ambiguity surface clearly shows an improved result. Note that between 17–18.5 min the two targets are too close to resolve in range.

2. CLEAN with two ship tracks

To demonstrate the CLEAN algorithm, we generated a two-target situation by superimposing the CSDM of the ship between 14–18 minutes elongated twice in time resulting in half the ship speed and reversed in time with 6 dB higher intensity than the original target. Then the stronger target moves from 4 km to 3.3 km between 15–23 min. The resulting two ship trajectories are shown as long-dashed lines in the top panel of Fig. 7.

Figure 7 shows the ambiguity surface for a two-target scenario: Bartlett (top), WNC (middle), and CLEAN map (bottom). The intensity of the second target moving from 4 km to 3.3 km with a half speed is 6 dB higher than the original ship track (from 3.4 km to 5.5 km). As expected, the WNC processor (middle) shows better performance than the conventional Bartlett processor (top), but the weaker ship track is hardly visible. The cleaning process applied to the WNC ambiguity surface clearly shows an improved result over the dirty map in the middle. Note that between 17–18.5 min the two targets are too close to resolve in range.

To investigate in detail the improvement of the CLEAN process associated with sidelobe control, we focus on the first minute starting at 15 min with six frames (10 s for each frame) in Fig. 7. Figure 8 shows the results of the residuals over the 4–6 CLEAN iterations. Blue and red vertical lines indicate the position of the strong and weak target in range, respectively. The original WNC ambiguity surface is shown by a thick solid line along with competing sidelobes, one of which corresponds to a weak target. Thick dashed lines indicate the residuals after a few iterations, whose peaks occur at the weaker target. Note that the sidelobes generated by the strong target are diminished over the iteration procedure, whereas the weaker target remains relatively unchanged.

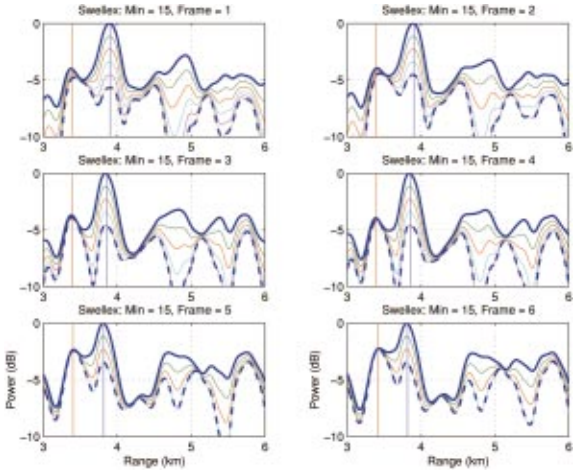


FIG. 8. The residual ambiguity surface over the process of the CLEAN algorithm iterations during a one-minute period starting 15 min with six frames. Blue and red vertical lines indicate the position of the strong and weak target, respectively. The original WNC ambiguity surface is shown by a thick solid line along with competing sidelobes. Thick dashed line indicates the residuals after a few iterations, whose peaks occurs at the weaker target. The sidelobes generated by the strong target are diminished over the iteration procedure, whereas the weaker target remains relatively unchanged.

V. SUMMARY

The CLEAN algorithm was introduced in radio astronomy to produce a high quality image of the sky by reducing sidelobe-induced artifacts. In this paper, the CLEAN concept is extended to matched field processing in the ocean. Numerical simulations and experimental data demonstrate that matched field processing combined with the CLEAN algorithm provides better performance, especially when a weak source is masked by sidelobes from a much stronger source.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research.

[1] A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, "An overview of matched field methods in ocean acoustics," IEEE J. Ocean. Eng. **18**, 401–424 (1993).
[2] J. A. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," Astron. Astrophys., Suppl. Ser. **15**, 417–426 (1974).
[3] J. Tsao and B. D. Steinberg, "Reduction of sidelobe and speckle artifacts in microwave imaging: The CLEAN technique," IEEE Trans. Antennas Propag. **20**, 543–556 (1988).
[4] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," IEEE Trans. Acoust., Speech, Signal Process. **10**, 1365–1376 (1987).
[5] W. A. Kuperman and F. Ingenito, "Spatial correlation of surface generated noise in a stratified ocean," J. Acoust. Soc. Am. **67**, 1988–1996 (1980).
[6] J. L. Yen, in *Array Signal Processing*, edited by S. Haykin (Prentice-Hall, Englewood Cliffs, NJ, 1985), pp. 293–350.
[7] *Synthesis Imaging in Radio Astronomy*, edited by G. B. Taylor, C. L. Carilis, and R. A. Perley, ASP Conference Series, Vol. 180, pp. 151–170, 1999.
[8] M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford, 1970).
[9] D. N. Forte and H. K. C. Yee, "A method of obtaining brightness distribution from long-baseline interferometry," Astron. Astrophys. **50**, 19–22 (1976).
[10] D. S. Briggs, Ph.D. thesis, New Mexico Institute of Mining and Technology, 1995.

[11] U. J. Schwarz, ''Mathematical-statistical description of the iterative beam removing technique (Method CLEAN),'' Astron. Astrophys. **65**, 345–356 (1978).
[12] U. J. Schwarz, ''The method 'CLEAN'—Use, misuse and variations,'' in *Image Formation from Coherence Functions in Astronomy*, edited by C. van Schooneveld (D. Reidel, New York, 1979), pp. 261–275.
[13] T. J. Cornwell, ''A method of stabilizing the CLEAN algorithm,'' Astron. Astrophys. **121**, 281–285 (1983).
[14] A. B. Baggeroer, W. A. Kuperman, and Henrik Schmit, ''Source localization in correlated noise as an optimum parameter estimation problem,'' J. Acoust. Soc. Am. **83**, 571–587 (1988).
[15] H. Cox, ''Resolving power and sensitivity to mismatch of optimum array processors,'' J. Acoust. Soc. Am. **54**, 771–785 (1973).
[16] N. O. Booth, ''SWellEx-96 preliminary data Report,'' NRaD/NCCOSC Code 881, San Diego, CA, 1996.
[17] W. S. Hodgkiss, J. J. Murray, N. O. Booth, and P. W. Schey, ''Near-surface source depth resolution in a shallow-water environment,'' J. Acoust. Soc. Am. **102**, 3171 (1997).

Absolute measurements of total target strength from reverberation in a cavity

David A. Demer
Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, California 92037

Stephane G. Conti
Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, California 92037, Universite Pierre et Marie Curie - Paris VI, 4 Place Jussieu, 75005 Paris, France, and Laboratoire Ondes et Acoustique, 10 Rue Vauquelin, 75005 Paris, France

Julien De Rosny
Marine Physical Laboratory, SIO/UCSD, 9500 Gilman Drive, La Jolla, California 92093-0205

Philippe Roux
Laboratoire Ondes et Acoustique, 10 Rue Vauquelin, 75005 Paris, France

(Received 30 October 2001; revised 9 October 2002; accepted 4 December 2002)

A new method was developed to acoustically measure the density and total scattering cross-section (σ_t) or total target strength [$\mathrm{TTS}=10\log_{10}(\sigma_t/4\pi)$] of objects in motion in a highly reflective cavity [J. De Rosny and P. Roux, J. Acoust. Soc. Am. **109**, 2587–2597 (2001)]. From an ensemble of pulse-echo recordings, the average contribution of the scatterer(s) to the reverberation within the cavity provides a measurement of the scattering mean free path. The latter was shown through theory and experiment to be proportional to the volume of the cavity and inversely proportional the product of the mean σ_t and number of scatterers. Here, the TTS measurement uncertainty is characterized using standard metal spheres as references. Theoretical TTS was calculated for multiple copper and tungsten carbide standard spheres (Cu: 60.0 30.05 and 23 mm and WC: 38.1 and 33.4 mm diameters, respectively), using well-described theory for scattering from elastic spheres and the optical theorem. Measurements of TTS were made over a wide bandwidth (30–120 kHz) and compared to their theoretical values. Measurements were made in a corrugated, cylindrical, galvanized-steel tank with 25 or 50 l of fresh water at a temperature of 21 ± 1 °C. The results indicate the method can provide TTS measurements that are accurate to at least 0.4 dB with an average precision of ±0.7 dB (95% confidence interval). Discussed are the requisite cavity volumes and signal-to-noise ratios for quality measurements of TTS, tank volume, and/or numerical abundance of mobile targets. Also discussed are multiple potential applications of this technique in bioacoustical oceanography. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1542648]

PACS numbers: 43.30.Xm, 43.30.Gv, 43.30.Sf, 43.80.Ev [DLB]

I. INTRODUCTION

Recently, a new method was invented to measure the number (n_s) of moving scatterers in a highly reverberant cavity (De Rosny, 2000). The theory was applied to acoustically counting fish in tanks (De Rosny and Roux, 2001). The method can also be used to estimate the volume of a cavity (v), or the total scattering cross-section (σ_t) of the scatterer(s) therein. Moreover, estimates of sound speed (c) and absorption cross-sections are also theoretically possible. Here, the accuracy and precision of the method for measuring σ_t is investigated using standard metal spheres as references.

In the new method, numerous pulses of sound ($i\in[1,N]$) are transmitted into a cavity having a static shape and volume. If the cavity hosts one or more mobile sound scatterers, the reverberation over time (t) resulting from the ith shot [$h^i(t)$] is comprised of echoes from (1) the motionless boundaries and (2) the moving object(s). For a single transmission, the two parts cannot be delineated. However, if the positions of the emitter and the receiver are fixed, the reverberation from the boundaries is coherent between records, while scatter from the moving objects is incoherent. By averaging N pulse-echo recordings (noted $\langle\,\rangle$), the parts of the signal due to the mobile scatterers are attenuated due to destructive interference, while that from the boundaries is reinforced. Following De Rosny and Roux (2001), the scattering mean free path (l_s, which characterizes σ_t, n_s, v, and c) can be estimated from the slope of the coherent sound intensity ($I_c=\langle h^i(t)\rangle^2$), divided by the incoherent intensity ($I_t=\langle h^i(t)^2\rangle$):

$$\frac{\langle h^i(t)\rangle^2}{\langle h^i(t)^2\rangle}\approx e^{-ct/l_s}. \tag{1}$$

This approximation does not depend on absorption in the medium nor on attenuation at the cavity interfaces, and is valid at time t when cavity echoes are present (De Rosny and Roux, 2001). It does require that the cavity conditions and boundaries are constant throughout the N recordings of $h^i(t)$.

TABLE I. Physical properties of copper (Cu) and tungsten carbide (WC) spheres that are relevant to their scattering cross-sections. Values for the material density (ρ_1) and longitudinal and transverse sound speeds (c_1 and c_2, respectively) for Cu and WC are from Foote and MacLennan (1984) and MacLennan and Dunn (1984), respectively.

	Cu	WC
ρ_1 (kg/m^3)	8947	14900
c_1 (m/s)	4760	6853
c_2 (m/s)	2288.5	4171

In practice, the cavity conditions are seldom static. For example, during the acquisition of N pulse-echo recordings, c may change appreciably due to temperature fluctuations. Therefore, l_s is better estimated from the slope ($d\ln(S)/dt$) of the average correlation of successive backscattered signals (De Rosny and Roux, 2001):

$$S(t)=\left(\frac{\langle h^i(t)h^{i+1}(t)\rangle}{\langle h^i(t)^2\rangle}\right)\approx e^{-ct/l_s} \tag{2}$$

$$\frac{d\ln(S)}{dt}\approx\frac{-cn_s\sigma_t}{\nu}. \tag{3}$$

The derivation of $S(t)$ is detailed in De Rosny and Roux (2001), and is therefore not repeated here. They showed that this estimator is totally dependent on the number of emitter-receiver positions in the average, the bandwidth of the emitted signal, and the ergodicity of the cavity (i.e., constancy of cavity conditions and boundaries). That is, the estimator is most robust for measurements that are averaged over multiple variations of the system (i.e., cavity+emitter +receiver). A practical approach to obtaining multiple system geometries is to use an omnidirectional emitter and receiver and a chirp transmission.

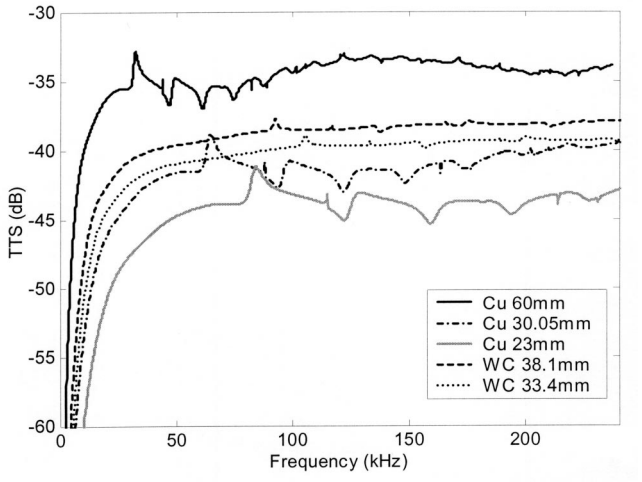


FIG. 1. Theoretical total target strengths [$\mathrm{TTS}=10\log_{10}(\sigma_t/4\pi)$] for standard spheres: electrolytic-grade copper (Cu) spheres (60-, 30.05-, 23-, and 13-mm diameters) and tungsten carbide (WC) with 6% cobalt binder (38.1- and 33.4-mm diameter). These curves were generated using the equations detailed in the Appendix, material properties from Table I, sound speed in water $c_3=1488$ m/s, and water density $\rho_3=1030$ kg/m^3. The spheres were suspended from loops of monofilament line affixed with epoxy into a small bore.

To determine the efficacy of this measurement technique, precision metal spheres were used as references. The theory of scattering from elastic spheres is well described (e.g., Faran, 1951; Hickling, 1962, 1964). MacLennan (1981) tabulated these equations for computing backscattering cross-sections of metal spheres with material density (ρ_1) and longitudinal and transverse sound speeds (c_1 and c_2, respectively). These equations are extended here, in the Appendix, to derive total scattering cross-sections (σ_t). Foote and MacLennan (1984) and MacLennan and Dunn (1984) provided accurate values of ρ_1, c_1, and c_2 for

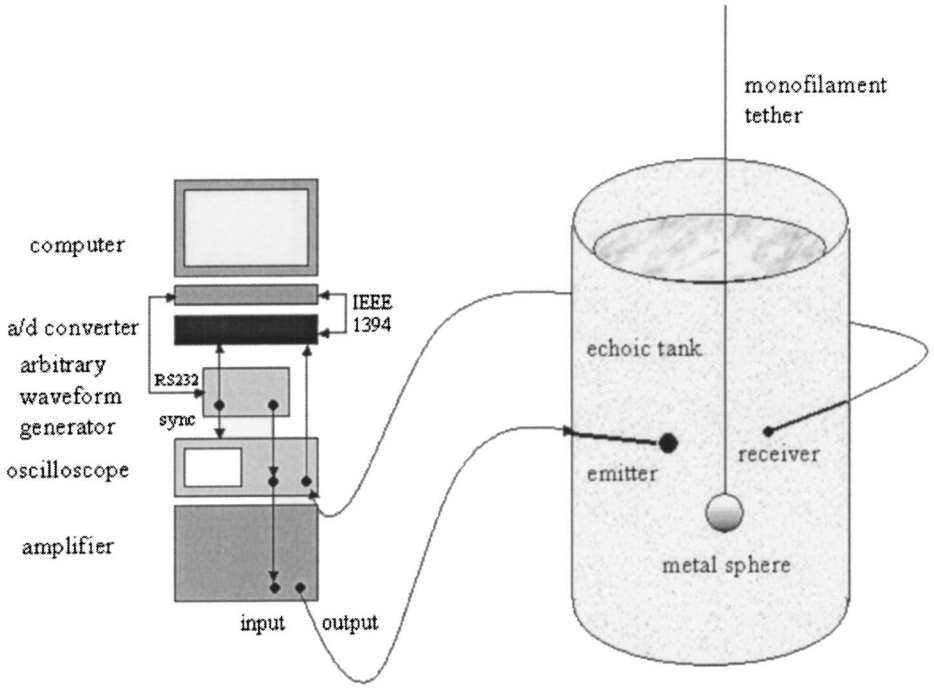
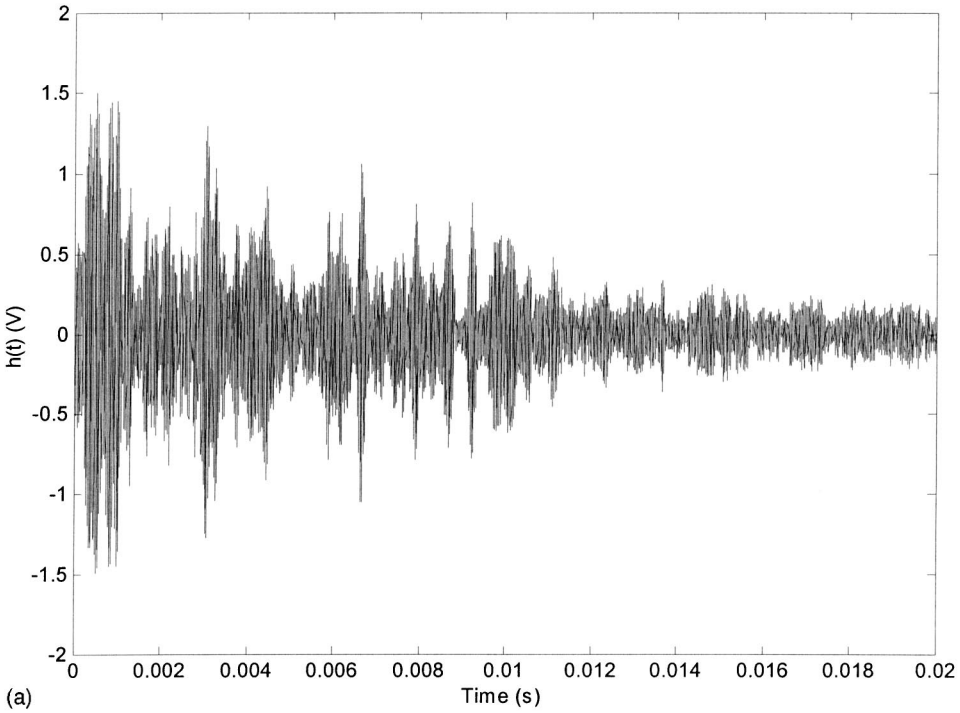


FIG. 2. Experimental apparatus. A corrugated, cylindrical, galvanized-steel tank was filled with fresh water. Chirp-pulses were created with an arbitrary waveform generator (Hewlett Packard 33120A), amplified by a broadband power amplifier (ENI 1140LA) and projected with an omnidirectional source (ITC 1042). Signals were received using an omnidirectional receiver (Reson 4013), digitized with a high-speed analog-to-digital converter (National Instruments Daqpad 6070E) and logged and processed on a laptop computer (Dell Inspiron 7500).



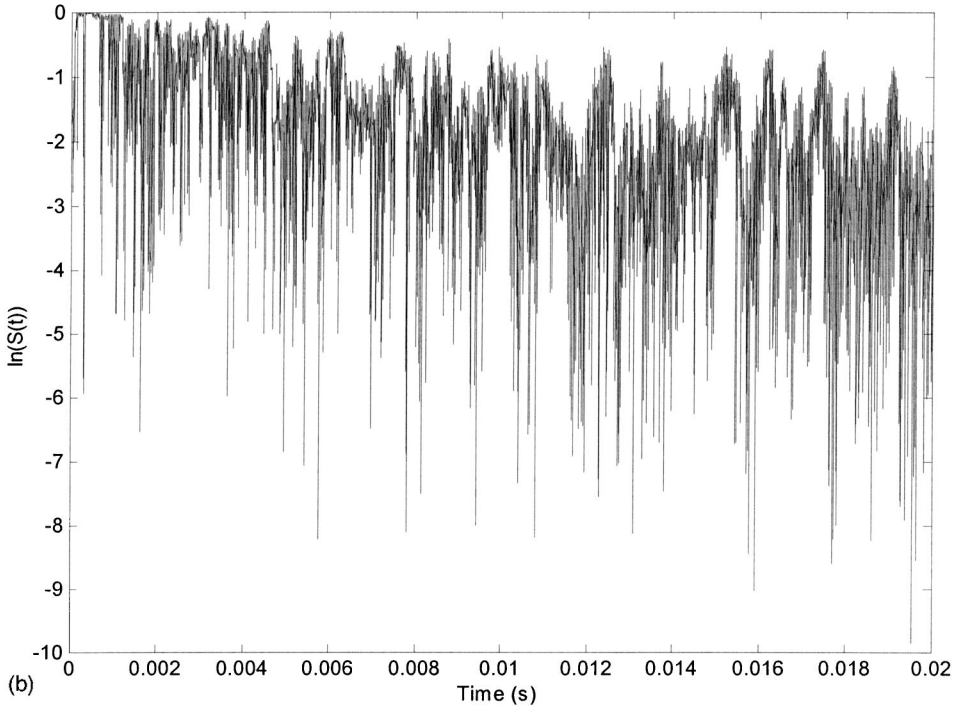


FIG. 3. Reverberation time-series $[h(t)]$ (a) and the natural logarithm of the estimator $S(t)$ (b) for a 60-mm-diam copper sphere (center frequency $=100$ kHz).

spheres made from copper (Cu) and tungsten carbide (WC) with 6% cobalt binder, respectively (Table I). Such spheres are routinely used as calibration references for scientific echosounders (e.g., Foote, 1982, 1990).

II. METHODS

Using the equations and material property values tabulated in the Appendix and Table I, respectively, theoretical σ_t were computed for five metal spheres (60, 30.05, and 23 mm Cu, and 38.1- and 33.4-mm-diam WC. The total target strength $[\mathrm{TTS}=10\log_{10}(\sigma_t/4\pi)]$ was also computed for each sphere (Fig. 1).

Measurements of TTS of metal spheres were made in August 2001 in the Advanced Survey Technologies Laboratory of the Southwest Fisheries Science Center. A corrugated, cylindrical, galvanized-steel tank, 54 cm in diameter and 67 cm tall was filled with 25 or 50 l of fresh water at a temperature $t_w=21\pm1$ °C (Fig. 2). Chirp-pulses of 500-μs duration,

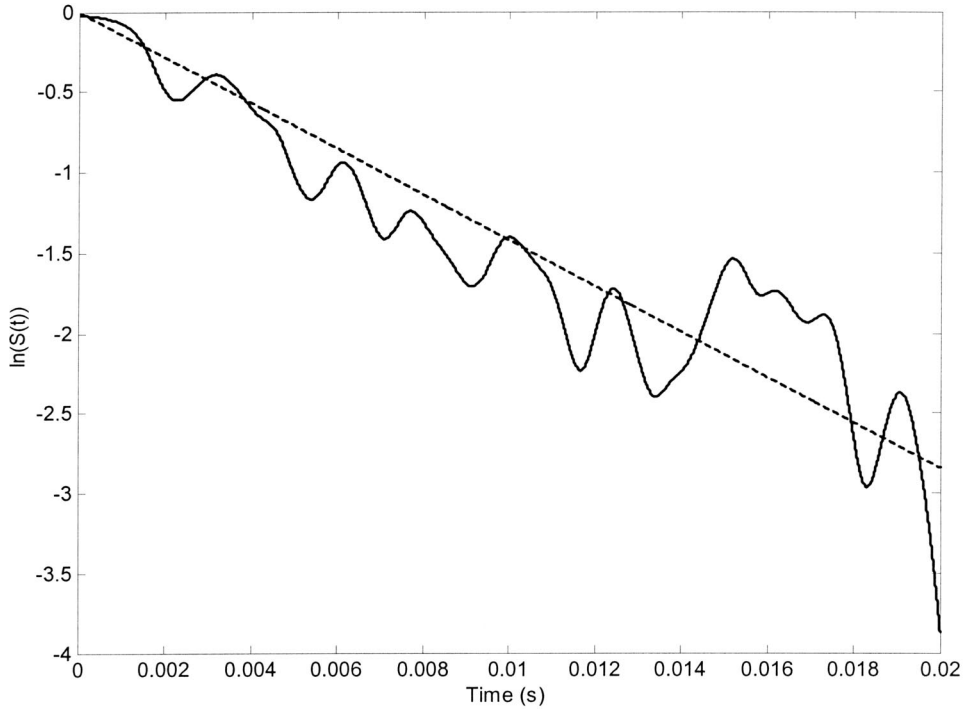


FIG. 4. The exponential decay of $S(t)$ was estimated for each 200-record ensemble by separately low-pass filtering the numerator and denominator of the function in the linear domain ($f_{\text{cutoff}} = 500$ Hz), transforming it to the logarithmic domain (—), and fitting a slope $[d \ln(S)/dt]$ in the least-squares sense, while requiring $2 \leqslant t \leqslant 15$ ms for the 50-l cavity and $2 \leqslant t \leqslant 10$ ms for the 25-l cavity, and $\ln(S) = 0$ at $t = 0$ (--). This example is from the 60-mm-diam copper sphere at a center frequency of 100 kHz.

2-Hz repetition rate, amplitudes of 400 or 170 $\text{mV}_{\text{p-p}}$, and center frequencies (f_c) from 30 to 50 kHz and 52 to 120 kHz, respectively, were created with an arbitrary waveform generator (Hewlett Packard 33120A). These 2-kHz bandwidth pulses were amplified 55 dB by a wide-bandwidth power amplifier (225 or 95.6 $\text{V}_{\text{p-p}}$; ENI 1140LA) and projected with a wide-bandwidth omnidirectional emitter (ITC 1042). The source was suspended 21.5 cm from the side of the tank and either 40 or 8 cm deep. Signals were received using a wide-bandwidth omnidirectional receiver (Reson 4013), placed 14 cm from the side of the tank and either 39.5 or 7.5 cm deep, depending on the water volume. The emitter and receiver were positioned towards the center of the tank so the observed sound pressure was not affected by the boundaries.

The precision metal spheres were placed in the tank one at a time. Continuously throughout the measurements, the spheres were laboriously moved by hand around the tank via the monofilament tether in a quasi-random three-dimensional motion. The speed of the manual motion was fast enough to

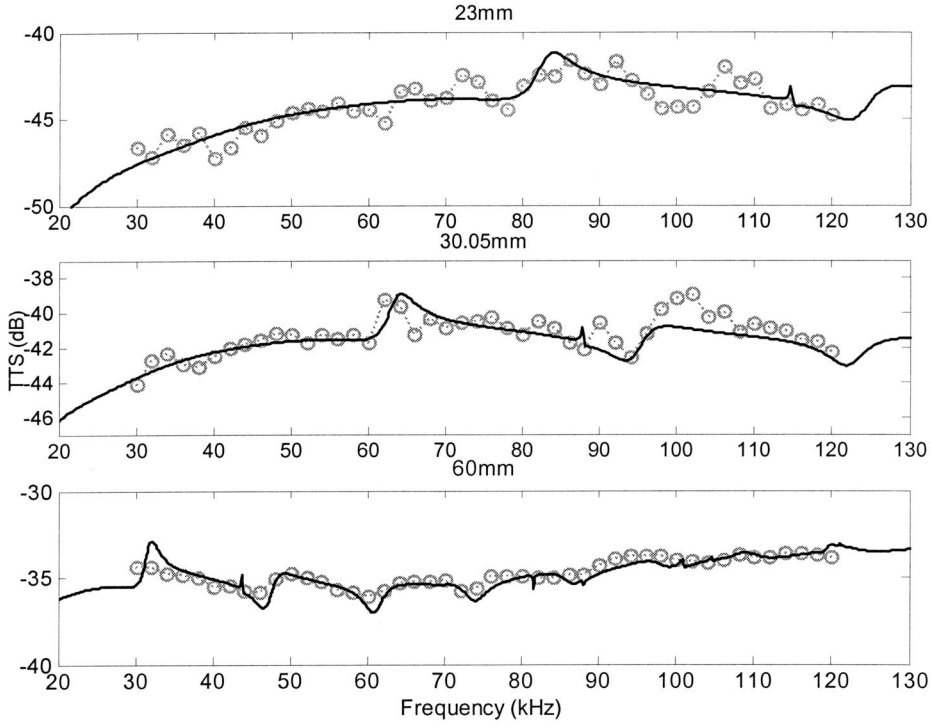


FIG. 5. Theoretical (−) and measured (-o-) TTS for 23-, 30.05-, and 60.0-mm-diam copper spheres. For the two smaller spheres, data were obtained in a single cavity, whereas the TTS measurements for the 60-mm-diam sphere were obtained with an average over three cavities (variations of receiver positions and tank volumes: 100 and 50 l). Differences between the measurements and theory are generally less than 1 dB.

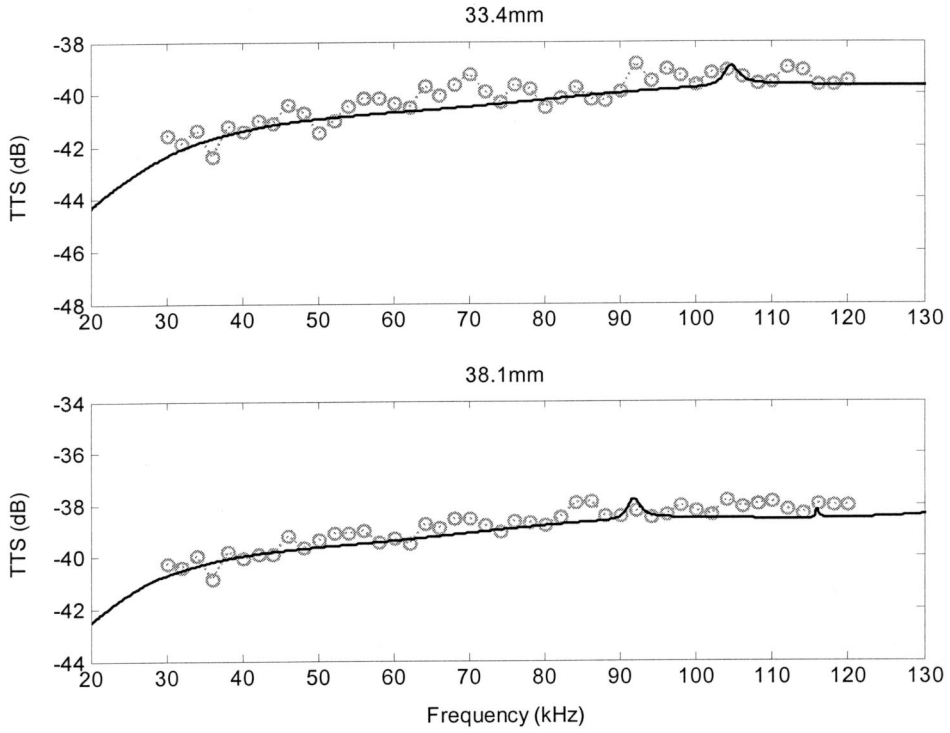


FIG. 6. Theoretical (−) and measured (-o-) TTS of 33.4- and 38.1-mm-diam tungsten carbide spheres (with 6% cobalt binder). Data were obtained for the 33.4-mm-diam sphere in a single cavity, whereas the TTS measurements for the 38.1-mm-diam sphere were obtained with an average over three cavities [variations of receiver positions and tank volumes (100 and 50 l)]. Again, differences between the measurements and theory are generally less than 1 dB.

maintain a very low correlation in reverberation between pings, yet slow enough that the water surface was not noticeably disturbed.

For each of 200 pings at each f_c, 32 ms of reverberation $[h^i(t)]$ were recorded using a 12-bit analog-to-digital converter (National Instruments Daqpad 6070E), sampling at 400 kHz (e.g., Fig. 3). The exponential decay of $S(t)$ was estimated for each 200-record ensemble by separately low-pass filtering the numerator and denominator of the function in the linear domain ($f_{\text{cutoff}}=500$ Hz), and fitting a slope $[d\ln(S)/dt]$ in the least-squares sense, while requiring $2\leqslant t \leqslant 15$ ms for the 50-l cavity and $2\leqslant t\leqslant 10$ ms for the 25-l cavity, and $\ln(S)=0$ at $t=0$ (Fig. 4). $S(t)$ was low-pass filtered before the ratio to average the nulls of the cavity. This procedure is useful for low frequencies, when a small number of modes are excited in the cavity. For higher frequencies, low-pass filtering $S(t)$ before or after the ratio provides the same results. From $d\ln(S)/dt$ an estimate of σ_t was made for each sphere and frequency:

$$\sigma_t \approx -\frac{\nu}{cn_s}\frac{d\ln(S)}{dt}. \tag{4}$$

Corresponding TTSs were then computed and compared to theory (Figs. 5 and 6). From the Appendix, theoretical values for σ_t (and TTS) can be equivalently computed by the discrete integration of the form function, an analytical integration of the Legendre polynomial, or by applying the optical theorem (Feenberg, 1932). All methods were tried, but for ease and speed of computations, the latter method was ultimately chosen.

To estimate the accuracy and precision of the method, the 60-mm-diam Cu and 38.1-mm-diam WC spheres were selected as references. For each of these two spheres, three estimates of TTS were obtained for each frequency. Mean values, standard deviations (sd), and 95% confidence intervals (± 2 sd) were computed for each frequency (Fig. 7).

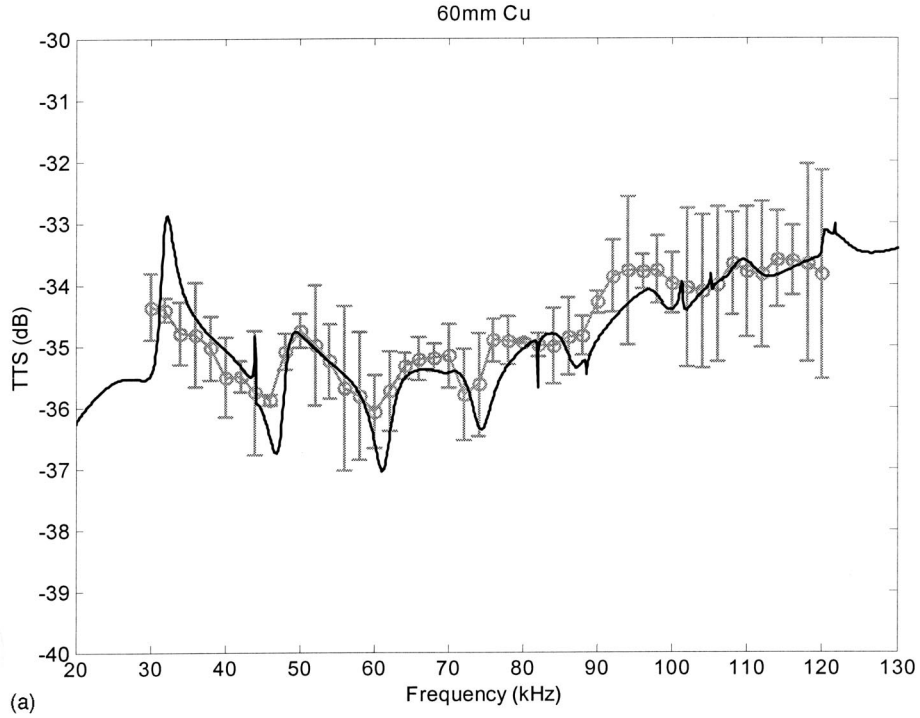
III. RESULTS

The wide-bandwidth measurements of TTS for the Cu and WC spheres were compared to their theoretical counterparts (Figs. 5 and 6, respectively). For all five spheres, there is a remarkable agreement between the empirical and theoretical data–mean discrepancies are only a few tenths of a decibel and the dynamics of the form function track closely. In some cases (e.g., the 30.05-mm-diam Cu sphere circa 62 kHz), slight discrepancies between the theoretical and empirical null positions may be due to a mismatch between the selected and actual sound speeds in the water [1488 vs 1485 ± 3 m/s, respectively; calculating $c(t_w)$ as in Lubbers and Graaf (1998)], or in c_1 or c_2.

Averaging three measurements at each frequency, TTS measurements were plotted with their 95% confidence intervals and compared to theory for both the 60 mm Cu and 38.1 mm WC spheres (Fig. 7). For the 60 mm Cu sphere, the mean difference between the theoretical and experimental TTS for each frequency is 0.46 dB with a sd of 0.34 dB. For the 38.1 mm WC sphere, the mean difference is 0.40 dB, with a sd of 0.36 dB.

IV. CONCLUSION

The new technique based on multiple scattering can be used to measure the TTS of a single scatterer moving in an echoic cavity, over a wide bandwidth and with low uncertainty. This study shows that for two spheres of different materials and sizes, the mean difference between the experimental estimates of TTS and theory is 0.4 dB with an average precision of ± 0.7 dB (95% confidence interval). Al-



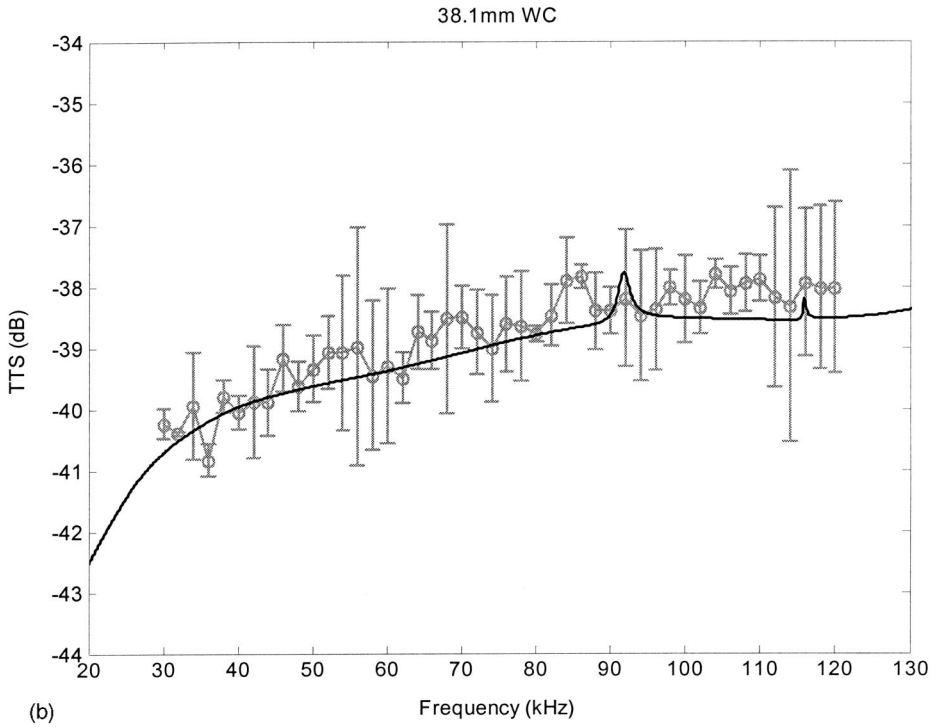


FIG. 7. TTS measurements (-o-) of the 60-mm-diam copper (a) and 38.1-mm-diam tungsten carbide spheres (b), and their 95% confidence intervals. The error bars (± 2 st. dev. $= \pm 0.67$ and ± 0.72 dB, respectively) generally encompass the theoretical TTS (—).

though this uncertainty is likely to be acceptable for most applications, there are some possibilities for improving upon these measurements.

Here, variables v and c were only estimated to within $\pm 0.5\%$ and 0.2%, respectively. The values of v and c could be more accurately estimated using traditional methods or by multi-dimensional application of the multi-scattering technique. That is, x-independent sets of measurements could be made and solved for x-unknown variables. The variance of the TTS measurements could be improved by a factor of $1/M$, by averaging M sets of σ_t measurements at each frequency. Moreover, the variance could be reduced by making physical changes to the cavity, its volume, and the positions of the emitter and receiver between measurement sets. Regarding the reference targets, there are also some minor uncertainties associated with the sizes, shapes, materials, densities, and c_1 or c_2 of the spheres (Foote and MacLennan, 1984).

It should also be mentioned that there are many requisite

considerations for making TTS measurements as good as the ones in this study. For example, to obtain a homogeneous sound field, a large number of modes must be excited in the cavity. Therefore, the characteristic size of the cavity, or smallest dimension, must be much greater than the wavelength ($\lambda = c/f$). As a guideline, $v \geqslant 100\lambda^3$ (e.g., $v \geqslant 12l$ at 30 kHz). Thus, the frequency range in these measurements (30 to 120 kHz) was rather limited by the frequency responses of the emitter and receiver. Additionally, v should not be too large compared to the total volume of the scatterers, and the reflectivity of the boundaries must be high, else the signal-to-noise ratio is insufficient. Many cavity materials were tested: fiberglass, high-density polyethylene, glass, and galvanized steel. Of these, the best reflectivity coefficients and thus the longest reverberation signals were obtained in the glass and galvanized steel tanks.

The TTSs of standard metal spheres were measured using the new multi-scattering method because the results could be confidently compared to primary standards. Although these measurements characterize the uncertainty in the multi-scattering method and constitute the first wide-bandwidth measurements of σ_t for metal spheres, other applications of the method are plentiful. De Rosny and Roux (2001) showed that the method could be used to accurately determine the number of fish in a tank. This could be useful in aquaculture. Of interest to the bioacoustical oceanographic community is the possibility for conveniently making wide-bandwidth measurements of total scattering cross-sections, resonant frequencies, and absorption cross-sections of oceanic sound scatterers. These data can be used to provide wide-bandwidth ''acoustical signatures'' for remotely identifying and classifying animals of a variety of species and sizes (Conti and Demer, submitted; and Demer and Conti, submitted).

Another interesting potential of this method is the characterization of σ_t as a function of animal size, shape, density, sound speed, and ambient pressure. Such data could be used for investigating both acoustical and biological properties of live animals and thus to improve the effectiveness of remote sensing tools such as echo-integration (e.g., Ehrenberg and Lytle, 1972), multiple-frequency target strength estimation and species delineation (Demer *et al.*, 1999), and bioacoustical absorption spectroscopy (e.g., Weston, 1967; Diachok *et al.*, 2001). Moreover, the data could provide much needed empirical validation of acoustical backscattering models (extended to σ_t) for a variety of fish (e.g., Love, 1977) and zooplankton (e.g., McGehee *et al.*, 1998; Demer and Conti, in press). While echo integration theory is based on backscattering, there are definable relationships (approximate if not analytically exact) between TTS and target strength for many scatterer types (morphologies, shapes, and orientation distributions), especially at low and high ka values.

ACKNOWLEDGMENTS

We are appreciative of William Kuperman, Director of Scripps Institution of Oceanography's Marine Physical Laboratory, and Rennie Holt, Director of the United States Antarctic Marine Living Resources Program, for co-sponsoring SC during his internship at the Southwest Fisheries Science Center. Thanks also to Charles Greenlaw, Ocean Sciences Group, BAE Systems; Andrew Brierley, Gatty Marine Laboratory, University of St. Andrews; and two anonymous reviewers for their helpful comments.

APPENDIX: THEORETICAL TOTAL SCATTERING CROSS-SECTION OF A SOLID ELASTIC SPHERE

The sound scatter from a solid elastic sphere in water was first computed by Faran (1951) and corrected by Hickling (1962, 1964). MacLennan (1981) tabulated these equations for computing the backscattering cross section of solid elastic spheres. The following is an extension of these equations, using Faran's symbols, to compute the total scattering cross-section.

In polar coordinates, for a plane wave of constant amplitude traveling in the z direction, the incident acoustic pressure at time t is

$$p_i(r,\theta,t) = p_0 \exp\{i(\omega t - k_3 r\cos\theta)\},$$

where k_3 is the wave number in the water and ω is the angular frequency. Interacting with a solid elastic sphere, the time-dependent scattered pressure is:

$$p_s(r,\theta,t) = p_{s0}(r,\theta)\exp(i\omega t),$$

where the forward- and backscattered pressures are defined at angles $\theta = 0$ and $\theta = \pi$, respectively. The scattered pressure can also be written as a sum of partial waves:

$$\begin{aligned} p_{s0}(r,\theta) &= -p_0 \sum_{n=0}^{\infty} (-i)^{n+1}(2n+1)\sin\eta_n \\ &\quad \times \exp(i\eta_n) h_n(k_3 r) P_n(\cos\theta) \\ &= -p_0 f(\theta, r), \end{aligned}$$

where $P_n(\cos\theta)$ is a Legendre polynomial, and $h_n(k_3 r)$ is the spherical Hankel function of the second kind, defined by

$$h_n(x) = j_n(x) - i n_n(x),$$

$$j_n(x) = J_{n+1/2}(x)\sqrt{\frac{\pi}{2x}}, \quad n_n(x) = N_{n+1/2}(x)\sqrt{\frac{\pi}{2x}},$$

where J and N are the Bessel functions of the first and second kind, respectively. The following equations determine η_n:

$$x_3 = k_3 a, \quad x_1 = \frac{x_3 c_3}{c_1}, \quad x_2 = \frac{x_3 c_3}{c_2},$$

$$A_2 = (n^2 + n - 2) j_n(x_2) + x_2^2 j_n''(x_2),$$

$$A_1 = 2n(n+1)[x_1 j_n'(x_1) - j_n(x_1)],$$

$$\alpha = 2\left(\frac{\rho_1}{\rho_3}\right)\left(\frac{c_2}{c_3}\right)^2, \quad \beta = \left(\frac{\rho_1}{\rho_3}\right)\left(\frac{c_1}{c_3}\right)^2 - \alpha,$$

$$\begin{aligned} B_2 &= A_2 x_1^2 [\beta j_n(x_1) - \alpha j_n''(x_1)] \\ &\quad - A_1 \alpha [j_n(x_2) - x_2 j_n'(x_2)], \end{aligned}$$

$$B_1 = x_3[A_2 x_1 j_n'(x_1) - A_1 j_n(x_2)],$$

$$\tan \eta_n = -\frac{[B_2 j_n'(x_3) - B_1 j_n(x_3)]}{[B_2 n_n'(x_3) - B_1 n_n(x_3)]},$$

where ρ_3 is the density of the water, c_3 is the sound speed in the water, a is the radius of the sphere, ρ_1 is the density of the sphere material, and c_1 and c_2 are the speeds of the longitudinal and transverse waves within the sphere, respectively (Table I). In the far field, at large range (r), or $k_3 r \gg 1$, the form function is

$$f(\theta) = \frac{1}{k_3} \sum_{n=0}^{\infty} (2n+1) \sin \eta_n \exp(i \eta_n) P_n(\cos \theta).$$

The total scattering cross section (σ_t) can be equivalently computed using three different methods: (1) a discrete integration of the form function, (2) an analytical integration of the Legendre polynomial, or (3) by application of the optical theorem (Feenberg, 1932). The discrete integration can be computed by

$$\sigma_t = \int_0^{2\pi} \int_0^{\pi} |f(\theta)|^2 d \cos \theta \, d\varphi$$

$$= -2\pi \int_0^{\pi} \sin \theta |f(\theta)|^2 \, d\theta.$$

Analytical integration of the Legendre polynomials is

$$\sigma_t = \int |f|^2 d\Omega = \int_0^{2\pi} \int_0^{\pi} \left[\sum_n \frac{1}{k_3} (2n+1) \sin \eta_n \times \exp(i \eta_n) P_n(\cos \theta) \right] \times \left[\sum_m \frac{1}{k_3} (2m+1) \sin \eta_m \times \exp(i \eta_m) P_m(\cos \theta) \right] d \cos \theta d\varphi$$

$$\sigma_t = \frac{2\pi}{k_3^2} \sum_{nm} \sin \eta_n \sin \eta_m \exp(i(\eta_n - \eta_m))(2n+1)(2m+1) \times \int_0^{\pi} P_n(\cos \theta) P_m(\cos \theta) d \cos \theta,$$

$$\sigma_t = \frac{2\pi}{k_3^2} \sum_n \sin^2 \eta_n \exp(i(\eta_n - \eta_n))(2n+1)^2 \frac{2}{2n+1},$$

$$\sigma_t = \frac{4\pi}{k_3^2} \sum_{n=0}^{\infty} (2n+1) \sin^2 \eta_n .$$

Finally, the conventional optical theorem involves the imaginary part (Im) of the forward-scattering form function:

$$\sigma_t = \frac{4\pi}{k_3} \mathrm{Im}(f(0)).$$

Applied to elastic spheres, the total scattering cross section is efficiently derived by

$$\sigma_t = \frac{4\pi}{k_3^2} \sum_{n=0}^{\infty} (2n+1) \sin \eta_n \, \mathrm{Im}(\exp(i \eta_n)).$$

Conti, S. G., and Demer, D. A. (**submitted**). "Wide-bandwidth acoustical characterization of anchovy and sardine from reverberation measurements in an echoic tank," ICES J. Mar. Sci.

Demer, D. A., and Conti, S. G. (**submitted**). "Broadbandwidth total target strength measurements of Antarctic krill (*Euphausia superba*) from reverberation in a cavity," ICES J. Mar. Sci.

Demer, D. A., and Conti, S. G. (**in press**). "Reconciling theoretical versus empirical target strengths of krill; effects of phase variability on the distorted wave Born approximation," ICES J. Mar. Sci.

Demer, D. A., Soule, M. A., and Hewitt, R. P. (**1999**). "A multiple-frequency method for potentially improving the accuracy and precision of *in situ* target strength measurements," J. Acoust. Soc. Am. **105**(4), 2359–2376.

De Rosny, J. (**2000**). "Milieux réverbérants et réversibilité," Thèse de Doctorat Université Paris VI.

De Rosny, J., and Roux, P. (**2001**). "Multiple scattering in a reflecting cavity: Application to fish counting in a tank," J. Acoust. Soc. Am. **109**, 2587–2597.

Diachok, O., Liorzou, B., and Scalabrin, C. (**2001**). "Estimation of the number density of fish from resonance absorptivity and echo sounder data," ICES J. Mar. Sci. **58**, 137–153.

Ehrenberg, J. E., and Lytle, D. W. (**1972**). "Acoustic Techniques for Estimating Fish Abundance," IEEE Trans. Geosci. Electron. **GE-10**, 138–145.

Faran, J. J. (**1951**). "Sound scattering by solid cylinders and spheres," J. Acoust. Soc. Am. **23**, 405–418.

Feenberg, E. (**1932**). "The Scattering of Slow Electrons by Neutral Atoms," Phys. Rev. **40**(1), 40–54.

Foote, K. G. (**1982**). "Optimizing copper spheres for precision calibration of hydroacoustic equipment," J. Acoust. Soc. Am. **71**, 742–747.

Foote, K. G. (**1990**). "Spheres for calibrating an eleven-frequency acoustic measurement system," J. Cons., Cons. Int. Explor. Mer **46**, 284–286.

Foote, K. G., and MacLennan, D. N. (**1984**). "Comparison of copper and tungsten carbide calibration spheres," J. Acoust. Soc. Am. **75**, 612–616.

Hickling, R. (**1962**). "Analysis of echoes from a solid elastic sphere in water," J. Acoust. Soc. Am. **34**, 1582–1592.

Hickling, R. (**1964**). "Analysis of echoes from a hollow metallic sphere in water," J. Acoust. Soc. Am. **36**, 1124–1137.

Love, R. H. (**1977**). "Target strength of an individual fish at any aspect," J. Acoust. Soc. Am. **62**, 1397–1403.

Lubbers, J., and Graaf, R. (**1998**). "A simple and accurate formula for the sound velocity in water," Ultrasound Med. Biol. **24**(7), 1065–1068.

MacLennan, D. N. (**1981**). "The theory of solid spheres as sonar calibration targets," Scottish Fisheries Research Report, No. 22.

MacLennan, D. N., and Dunn, J. R. (**1984**). "Estimation of sound velocities from resonance measurements on tungsten carbide calibration spheres," J. Sound Vib. **97**, 321–331.

McGehee, D. E., O'Driscoll, R. L., and Martin-Traykovski, L. V. (**1998**). "Effects of orientation on acoustic scattering from Antarctic krill at 120 kHz," Deep-Sea Res., Part II **45**, 1273–1294.

Weston, D. E. (**1967**). "Sound propagation in the presence of bladder fish," in *Underwater Acoustics*, edited by V. M. Albers (Plenum, New York), Vol. 2, pp. 55–88.

Analysis of a compliantly suspended acoustic velocity sensor

James A. McConnell[a)]
Acoustech Corporation, P.O. Box 139, State College, Pennsylvania 16804

(Received 27 July 2001; revised 8 October 2002; accepted 4 December 2002)

The dynamics of a compliantly suspended acoustic velocity sensor having a spherical geometry are analyzed using theory and experiment. The analysis starts with a review of the motion associated with an unconstrained solid sphere when subjected to an acoustic plane wave in an unbounded inviscid fluid medium. The theory is then modified to account for the inclusion of an inertial sensor and an external suspension system. Accordingly, the open-circuit receiving response of a geophone-based and accelerometer-based device is derived. Density variations associated with the sphere and the surrounding fluid medium are assessed along with the effects fluid viscosity. Wave effects in the sphere and the suspension system are also analyzed. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1542646]

PACS numbers: 43.30.Yj, 43.58.Fm, 43.30.Xm, 43.38.Ar [DLB]

I. INTRODUCTION

In practice, underwater acoustic particle velocity measurements are performed using either inertial sensors or gradient sensors. Inertial sensors typically rely on the concept of imbedding a moving coil geophone or piezoelectric accelerometer in a small rigid body and recording the output voltage that results when the body is set into motion. This concept is largely based on the theory derived for the response of an unconstrained rigid sphere when subjected to an acoustic plane wave in an inviscid unbounded fluid medium. The analysis shows that if the sphere is small compared to an acoustic wavelength, its velocity amplitude is related to that of the acoustic wave by[1–5]

$$\frac{V_s}{V_o}=\frac{3\rho_o}{2\rho_s+\rho_o}\equiv\gamma, \tag{1}$$

where V_s is the velocity amplitude of the sphere, V_o is the velocity amplitude of the acoustic wave (e.g., the acoustic particle velocity), ρ_o is the density of the fluid medium, and ρ_s is the density of the sphere. It can be concluded from Eq. (1) that a positively buoyant sphere responds with an amplitude that is greater than the acoustic wave, a neutrally buoyant sphere responds with an amplitude that is the same as the acoustic wave, and a negatively buoyant sphere responds with an amplitude that is less than the acoustic wave. Also, since the sphere is unconstrained, there are no restoring forces; hence, the phase of the sphere is the same as the acoustic wave. This analysis indicates that a neutrally buoyant sphere moves in concert with the acoustic wave, and by extension, serves as the basis for many devices that are used to measure the acoustic particle velocity in water. Some relevant papers that illustrate this concept are by Leslie *et al.*,[2] Gabrielson *et al.*,[6] Moffett *et al.*,[7] and Bastyr *et al.*[8] Others can be found in the proceedings from two fairly recent conferences.[9,10]

In contrast to inertial sensors, gradient sensors must be fixed in space and are designed to have a voltage output that is proportional to the pressure gradient. In this case the particle velocity is inferred from the linearized Euler equation, e.g., $\nabla p=-\rho_o a$, where p is the acoustic pressure and a is the acoustic particle acceleration.

Gradient sensors generally employ two different design concepts. One concept considers separating two omnidirectional pressure sensors by a small distance and wiring the outputs out of phase so that the signals are subtracted. This concept serves as the basis for the two-microphone technique commonly used for airborne intensity and impedance measurements.[11,12] It is noted, however, that this technique is only viable as long as the absolute sensitivities of the sensors are matched (or can be matched through calibration) and the separation distance is small compared to a wavelength. The latter requirement is mandated to minimize the error associated with the finite-difference approximation to the Euler equation.[13]

Other pressure gradient sensors rely on the principle of subjecting both sides of a baffled flexural transducer (e.g., a bimorph) to a sound wave so that the net voltage output corresponds to the difference in pressure across the flexural element. Pressure gradient sensors of this type are described in detail by Woollett[14] and Bobber.[15]

The focus of this paper concerns the analysis of Eq. (1) when the sphere contains an accelerometer and an external suspension system. Simply stated, the addition of these two parameters changes the dynamics of the system significantly and requires a thorough evaluation of the underlying physics so that the development of such transducers leads to practical devices that have good fidelity in the preferred measurement band. Of particular interest is the impact that external suspension systems has on the fidelity of the device. This is important because, in practice, inertial sensors have to be suspended from some host platform; therefore, it is critical that the suspension system and/or the host platform not contaminate the measurement. The effect of varying the fluid and sphere densities is also investigated along with the effects of fluid viscosity. The analysis is carried out on a theoretical basis and is compared to results obtained by a series of experiments.

[a)]Electronic mail: mcconnell@acoustechcorporation.com

II. THEORY OF SPHERICAL VELOCITY SENSORS

A. Unconstrained sphere

In the previous section, Eq. (1) showed that the motion of a small unconstrained sphere in an acoustic plane-wave field is purely a function of density. The formulation is specific to the case of the sphere and is based on the following equation of motion, which equates the force on the sphere to the reaction force of the fluid:[1–5]

$$j\omega(m_s+m_i)V_s e^{j(\omega t-kx)}=j\omega(m_o+m_i)V_o e^{j(\omega t-kx)}, \quad (2)$$

where m_s is the mass of the sphere, m_i is the induced mass (i.e., the reactive term associated with the acoustic radiation impedance, see Mofrey[16] for a more detailed definition) that is created as a result of the sphere translating in the fluid, and m_o is the mass of fluid that is displaced by the sphere. Note that the sphere is required to be small compared to a wavelength to preclude scattering effects.[2] Also, it is noted that the $e^{j(\omega t-kx)}$ dependence is used and will be suppressed in subsequent formulations. Nevertheless, Eq. (2) can be rewritten as

$$\frac{V_s}{V_o}=\frac{m_o+m_i}{m_s+m_i}, \quad (3)$$

and is valid for an arbitrary-shaped body whose dimensions are small compared to a wavelength. Now, for the case of the sphere, the induced mass can be computed analytically in closed form and is equal to $m_o/2$, or $2\pi a^3\rho_o/3$, where a is the radius of the sphere.[1–5,17] From this formulation, the result shown in Eq. (1) is easily obtained.

Moreover, if the geometry of the sensor is that of a thin disk of radius a and thickness t, then the induced mass is determined in closed form[1,17] to be $8a^3\rho_o/3$, which leads to

$$\frac{V_d}{V_o}=\frac{(1+\alpha)\rho_o}{\alpha\rho_d+\rho_o}, \quad (4)$$

where V_d is the velocity amplitude of the disk translating along its axis, $\alpha=3\pi t/8a$, and ρ_d is the density of the disk. For the case of neutral buoyancy, the disk moves in concert with the fluid, as does the sphere; however, departures from neutral buoyancy lead to formulations of Eq. (4) which are a function of both density and geometry (and not density alone). Furthermore, for the case of a right circular cylinder translating along its axis, there is no closed-form solution for the induced mass. This quantity must be calculated numerically or by experiment. Therefore, Eq. (3) cannot be solved in closed form and may present a relatively small challenge to the transducer designer when non-neutrally buoyant sensors having a cylindrical geometry are considered. Nevertheless, Blevins[17] presents a detailed summary of the induced mass for numerous body geometries and is of significant value to the current research.

When fluid viscosity is considered, the force balance in Eq. (2) must be modified to include the terms that are associated with the induced drag that results when the body is set into motion. Under the assumption of a low Reynolds number, Re, Fuchs[5] and Temkin[18] derived the equation of motion using the Stoke's drag on a sphere in a viscous oscillatory fluid medium, namely

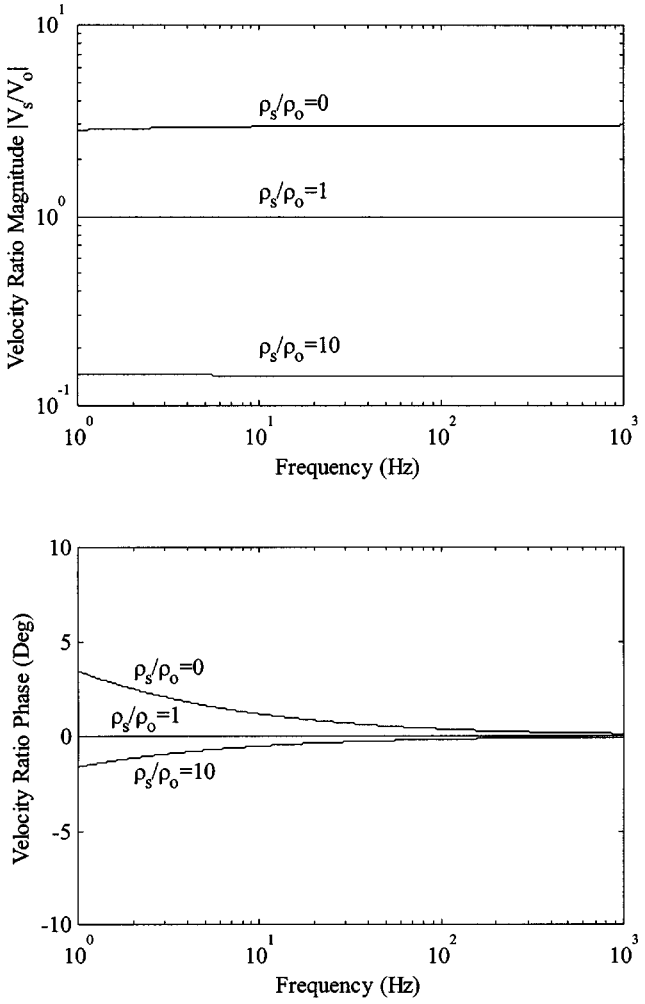


FIG. 1. Magnitude and phase response of an unconstrained sphere when subjected to an acoustic plane wave in an unbounded viscous fluid medium. The sphere has a diameter of 5.08 cm and is immersed in water at 20 °C.

$$j\omega(m_s+m_i)V_s=j\omega(m_o+m_i)V_o-\tfrac{9}{4}m_o\omega\beta(1+\beta)\times(V_s-V_o)-j\tfrac{9}{4}m_o\omega\beta(V_s-V_o), \quad (5)$$

where $\beta=a^{-1}(2\nu/\omega)^{1/2}$ and ν is the kinematic viscosity of the fluid. The term β is the viscous boundary layer thickness normalized by the radius of the sphere and is considered in this analysis to be arbitrary. Equation (5) can be rewritten as

$$\frac{V_s}{V_o}=\frac{3\rho_o+9\rho_o\beta[1-j(1+\beta)]/2}{2\rho_s+\rho_o+9\rho_o\beta[1-j(1+\beta)]/2}, \quad (6)$$

and is valid for $\mathrm{Re}\leqslant 3$, where $\mathrm{Re}=2V_o a/\nu$. The restriction on Re is invariably met for wave propagation that conforms to the equations of linear acoustics (e.g., when $V_o/c\ll 1$, where c is the sound speed in the fluid).

Inspection of Eq. (6) shows that it has the familiar form of Eq. (1) with the added terms to account for the viscous drag. Figure 1 presents the result of plotting Eq. (6) versus frequency for the case of ρ_s/ρ_o equal to 0, 1, and 10. For this analysis, the sphere has a diameter of 5.08 cm and is immersed in water at 20 °C, e.g., $\nu=1.005\times10^{-6}$ m^2/s, such that $\beta=0.01$ at $f=1$ Hz. Figure 1 shows that the response of a neutrally buoyant sphere is identical to that of the acoustic wave and is frequency independent. However, the response

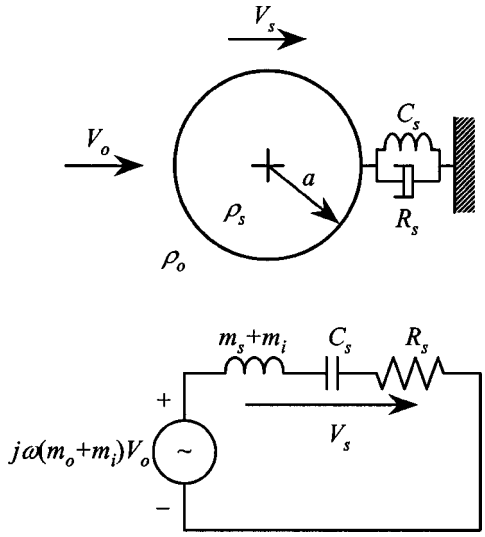


FIG. 2. Schematic of a compliantly suspended sphere undergoing rigid body translation when subjected to an acoustic plane wave. The equivalent mechanical impedance circuit is also shown.

of non-neutrally buoyant spheres varies with frequency and asymptotically approaches the values predicted by Eq. (1) as the frequency is increased. This result is particularly evident with respect to the phase response of the sphere.

The physical interpretation of this result indicates that for a given fluid viscosity, increasing the size of the sphere preserves the response at low frequencies, but as a consequence lowers the inception point where scattering effects cannot be ignored. It can be inferred from Eq. (6) that viscous effects can be neglected provided that $\beta \ll 1$.

B. Compliantly suspended sphere

The dynamics of the problem are somewhat different when the sphere is constrained by a compliant suspension system. Figure 2 shows a schematic of a suspended sphere along with a lumped parameter circuit model of the system. In Fig. 2, the spring has compliance C_s, the damper has resistance R_s, and the fluid is assumed to be inviscid. The suspension is grounded to a rigid structure and is not exposed to the acoustic wave. The equation of motion can be inferred from the circuit and ultimately leads to the following formulation:

$$\frac{V_s}{V_o} = \frac{3\rho_o}{2\rho_s + \rho_o} \cdot \left[1 - 2j\zeta_s \frac{\omega_s}{\omega} - \frac{\omega_s^2}{\omega^2} \right]^{-1}, \tag{7}$$

where ω_s is the resonance frequency and ζ_s is the damping factor. These parameters are further defined as $\omega_s = [(m_s + m_i)C_s]^{-1/2}$ and $\zeta_s = R_s/2(m_s + m_i)\omega_s$.

Figure 3 presents the result of plotting Eq. (7) versus dimensionless frequency ω/ω_s for damping factors that span two orders of magnitude. The analysis is specific to the case of neutral buoyancy. It can be inferred from the figure that about a decade above resonance, the velocity amplitude of the sphere is identical to that of the acoustic wave. The phase response, however, depends on the level of damping present in the system, and in particular, only lightly damped spheres exhibit little or no phase shift with respect to that of the acoustic wave. A conservative interpretation of these results

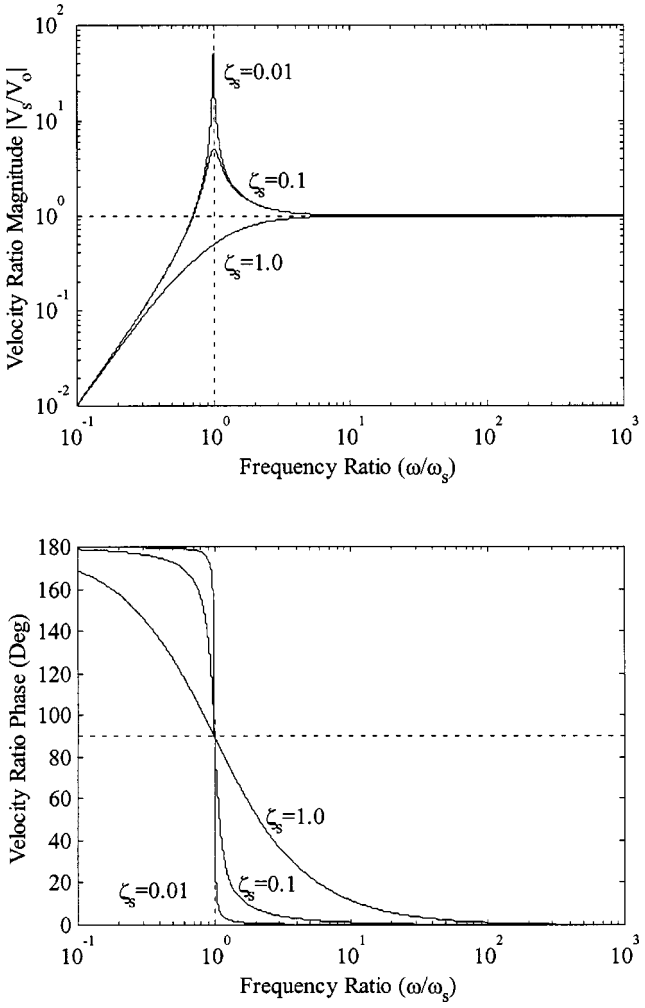


FIG. 3. Magnitude and phase response of a compliantly suspended sphere undergoing rigid body translation when subjected to an acoustic plane wave in an unbounded inviscid fluid medium.

suggests that the motion of a lightly damped sphere can be described by Eq. (1) at about a decade above resonance.

A special case of the foregoing analysis concerns the dynamics of a rigid sphere whose exterior has been coated with an elastic solid, which in turn is locally constrained. The coating is basically a compliant spherical shell of inner and outer radii a and a_c, and density ρ_c. Now, there is more than one approach to analyze the dynamics of such a system. The concept considered here treats the coating to have discrete properties of mass, compliance, and damping such that the coated sphere can be modeled as a damped harmonic oscillator having two degrees of freedom. This is shown in Fig. 4, which presents a lumped parameter model of the system. In Fig. 4, the coating can be regarded as a compliantly suspended rigid spherical shell that is set into motion by the passage of an acoustic wave. The internal sphere moves relative to the coating by way of a spring and damper. The coating (e.g., the external sphere) has mass m_c, compliance C_c, and resistance R_c. The velocity amplitude of the coating is denoted as V_c.

Analysis of the circuit shown in Fig. 4 indicates that the

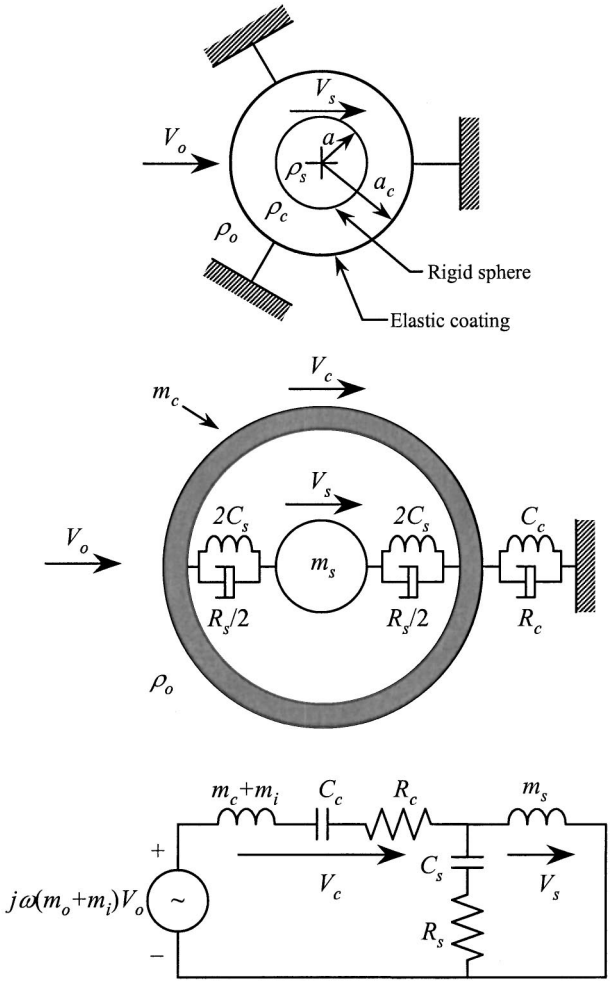


FIG. 4. Schematic of an elastically coated sphere undergoing rigid body translation when subjected to an acoustic plane wave. The lumped parameter representation is also shown along with the equivalent mechanical impedance circuit.

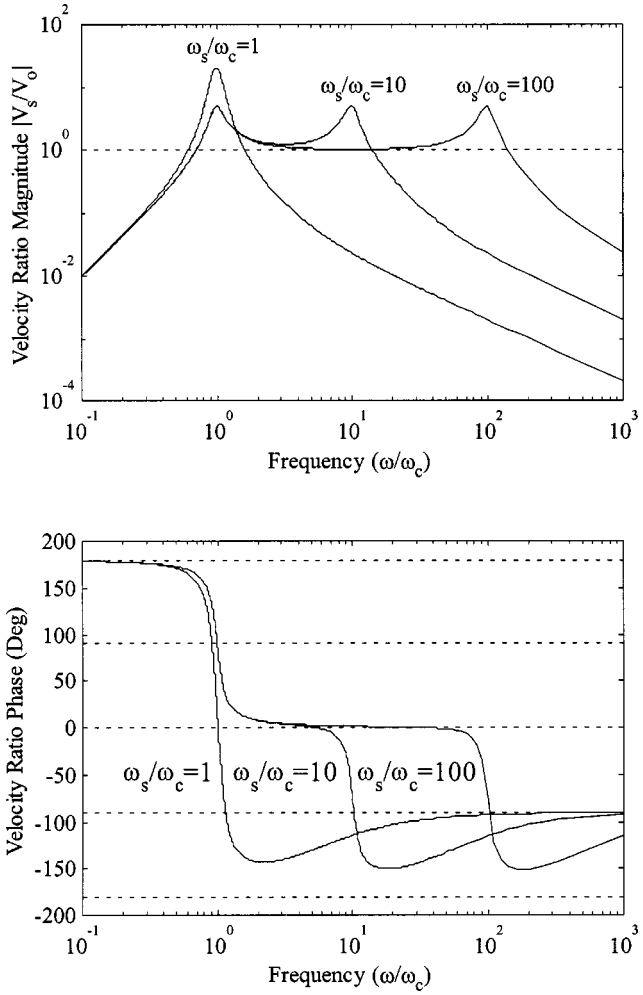


FIG. 5. Magnitude and phase response of an elastically coated sphere when subjected to an acoustic plane wave in an unbounded inviscid fluid medium.

velocity of the internal sphere is related to that of the acoustic wave by

$$V_s/V_o = Z(\omega), \tag{8}$$

where

$$Z(\omega) = -\frac{m_o+m_i}{m_c+m_i}\left(2j\zeta_c\frac{\omega_c}{\omega}+\frac{\omega_c^2}{\omega^2}\right)\cdot\left[\left(1-2j\zeta_s\frac{\omega_s}{\omega}-\frac{\omega_s^2}{\omega^2}\right)\times\left(1-2j\zeta_c\frac{\omega_c}{\omega}-\frac{\omega_c^2}{\omega^2}\right)-\frac{m_s}{m_c+m_i}\left(2j\zeta_c\frac{\omega_c}{\omega}+\frac{\omega_c^2}{\omega^2}\right)\right]^{-1}. \tag{9}$$

In Eq. (9), m_o and m_i are the displaced and induced fluid masses based on the outer radius of the coating a_c, respectively. Also, ω_c and ζ_c are the resonance frequency and damping factor associated with the coating, and ω_s and ζ_s are the resonance frequency and damping factor associated with the internal sphere. These parameters are further defined as $\omega_c=[(m_c+m_i)C_c]^{-1/2}$, $\zeta_c=R_c/2(m_c+m_i)\omega_c$, $\omega_s=(m_sC_s)^{-1/2}$, and $\zeta_s=R_s/2m_s\omega_s$.

Figure 5 presents the velocity ratio V_s/V_o computed from Eqs. (8) and (9) as a function of the resonance frequency ratio ω_s/ω_c. For this analysis, the mass ratio $(m_o+m_i)/(m_c+m_i)$ is arbitrarily set to unity to normalize the magnitude response. Under the assumption that the radius of the internal sphere is small compared to the thickness of the elastomeric coating (e.g., $a_c-a\gg a$) it can be assumed that mass ratio $m_s/(m_c+m_i)$ is very small and is arbitrarily set to be 0.01. Since the coating is elastomeric, such as polyurethane having a durometer of nominally 40 on the Shore A scale, the system compliance and damping will be moderately high; hence, the damping factors ζ_c and ζ_s are set to 0.1. It can be concluded from the figure that the response of the internal sphere conforms to that of the acoustic wave for frequencies that lie between the resonances of the system. In this region, $\omega/\omega_c\gg1$ and $\omega/\omega_s\ll1$; therefore, the velocity ratio tends to

$$\frac{V_s}{V_o}\cong\frac{m_o+m_i}{m_c+m_s+m_i}, \tag{10}$$

and conforms to Eq. (1) when the average density of the coated sphere is considered [e.g., $\rho_s=3(m_c+m_s)/4\pi a_c^3$]. The physical interpretation of this result is that in the passband between the resonances in the system, the internal sphere and coating move together as an unconstrained rigid

body. It should be stated, however, that a large separation between the resonances may be hard to realize in practice because the compliances C_c and C_s are likely to be the same, or at least very close. In this case, the coated sphere could contain an external suspension system in lieu of being locally constrained. If the compliance of the external suspension is much greater than C_c, then the low-frequency resonance could be reduced to maximize the bandwidth where the entire assembly moves as a rigid body.

For the idealized case of when the coating is more or less acoustically transparent, it can be regarded as a fluid-like body that confines the internal sphere. In this case, the induced mass of the internal sphere must be accounted for in Eqs. (8)–(10). Blevins[17] gives this relation to be

$$m_i = \frac{2}{3}\pi a^3 \rho_c \frac{1+2(a/a_c)^3}{1-(a/a_c)^3}. \tag{11}$$

Note that the foregoing analysis is considered to be a back-of-the-envelope estimate of the physics that governs the dynamics of an elastically coated sphere. The lumped parameter model neglects the contribution from elastic waves (e.g., both compressional and shear) that propagate through the coating and interact with the internal sphere. Keltie *et al.*[10] have analyzed the dynamics of an elastically coated sphere in a plane-wave field and include the effects of compressional and shear waves in the coating along with varying the density of the internal sphere with respect to that of the coating.

C. Compliantly suspended sphere with inertial sensor

The dynamics of the system are further complicated when an inertial sensor such as a piezoelectric accelerometer or moving coil geophone is imbedded in the sphere. In either case, the system can be modeled with lumped parameters to compute the open-circuit receiving response of the transducer. Figures 6 and 7 present schematics of acceleration-based and velocity-based spherical transducers along with their corresponding lumped parameter circuit models. As before, the spheres contain an external suspension system and are subjected to an acoustic plane wave in an inviscid unbounded fluid medium.

In Fig. 6, m_t, C_t, and R_t are the mechanical mass, compliance, and damping of the accelerometer, respectively. The transduction constant is denoted as ϕ and C_o is the clamped capacitance. The parameters A_o, A_s, and A_t, correspond to the acceleration amplitudes of the in-coming wave, sphere, and proof mass (e.g., m_t), respectively. Similarly, in Fig. 7, m_t, C_t, and R_t are the mechanical mass, compliance, and damping of the geophone, respectively. The transduction constant is denoted as Bl, R_e is the coil resistance, and L_e is the coil inductance. The parameter V_t corresponds to the velocity amplitude of the geophone proof mass (e.g., m_t). The open-circuit output voltage for each transducer is denoted as e_{oc}.

The open-circuit receiving response of the spherical accelerometer is determined from Fig. 6 to be $e_{oc}/A_o = \phi Z(\omega)/(j\omega)^2 C_o$. Likewise, from Fig. 7, the receiving response of the spherical geophone is determined to be $e_{oc}/V_o = BlZ(\omega)$. For both of these expressions, the term $Z(\omega)$ is defined as

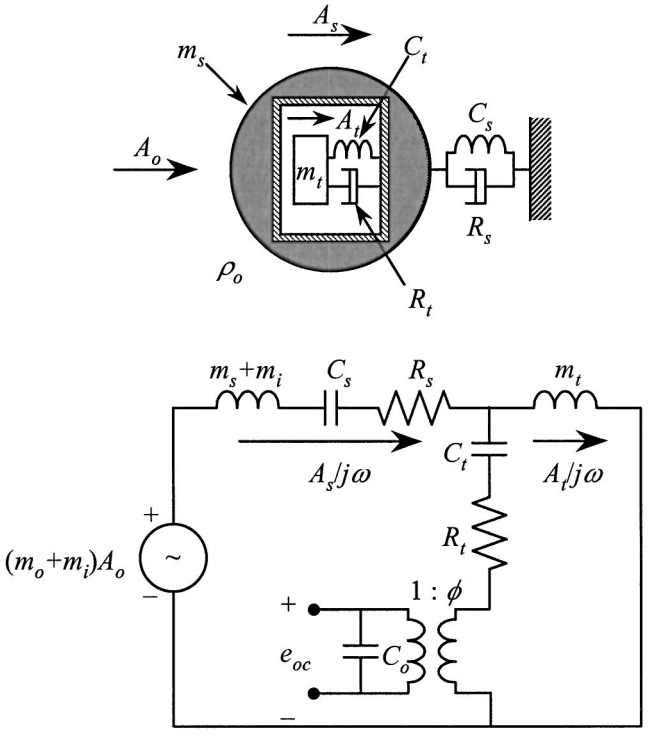


FIG. 6. Schematic of an acceleration-based spherical sensor with external suspension system. Sphere is undergoing rigid body translation as a result of acoustic plane-wave excitation. The equivalent electro-mechanical impedance circuit is also shown. The electrical and mechanical domains are coupled via a transformer having a transduction constant ϕ.

$$Z(\omega) = \frac{3\rho_o}{2\rho_s+\rho_o} \cdot \left[\left(1 - 2j\zeta_s \frac{\omega_s}{\omega} - \frac{\omega_s^2}{\omega^2} \right) \left(1 - 2j\zeta_t \frac{\omega_t}{\omega} - \frac{\omega_t^2}{\omega^2} \right) - \frac{m_t}{m_s+m_i} \left(2j\zeta_t \frac{\omega_t}{\omega} + \frac{\omega_t^2}{\omega^2} \right) \right]^{-1}, \tag{12}$$

where ω_s and ζ_s are the resonance frequency and damping factor associated with the sphere, and ω_t and ζ_t are the resonance frequency and damping factor associated with the inertial transducer inside the sphere. These parameters are further defined as $\omega_s = [(m_s+m_i)C_s]^{-1/2}$, $\zeta_s = R_s/2(m_s+m_i)\omega_s$, $\omega_t = (m_t C_{\text{eqv}})^{-1/2}$, and $\zeta_t = R_t/2m_t\omega_t$, where $C_{\text{eqv}} = C_t$ for the case of a geophone, and $C_{\text{eqv}} = C_o C_t/(C_o + \phi^2 C_t)$ for the case of an accelerometer. Note that the output voltage for either transducer corresponds to the relative motion between the sphere and the proof mass, and not the absolute motion of the sphere itself.

To assess the significance of these results, the receiving response of both transducers is plotted versus frequency $f = \omega/2\pi$. It is assumed that each sphere is lightly damped and exhibits a low resonance frequency such that $f_s = \omega_s/2\pi = 5$ Hz and $\zeta_s = 0.01$. It is also assumed that the resonance frequencies of the accelerometer and geophone are $f_t = \omega_t/2\pi = 25$ kHz and $f_t = 30$ Hz, respectively, and each transducer has a damping factor of $\zeta_t = 0.1$. Both spheres are assumed to be neutrally buoyant and the mass ratio $m_t/(m_s+m_i)$ is taken to be 0.01 (e.g., the proof mass of either

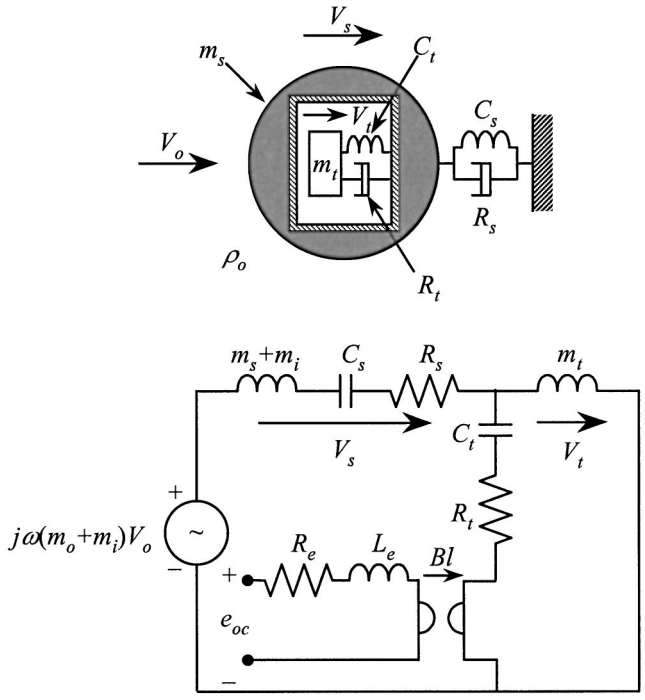


FIG. 7. Schematic of a velocity-based spherical sensor with external suspension system. Sphere is undergoing rigid body translation as a result of acoustic plane-wave excitation. The equivalent electro-mechanical impedance circuit is also shown. The electrical and mechanical domains are coupled via a gyrator having a transduction constant *Bl*.

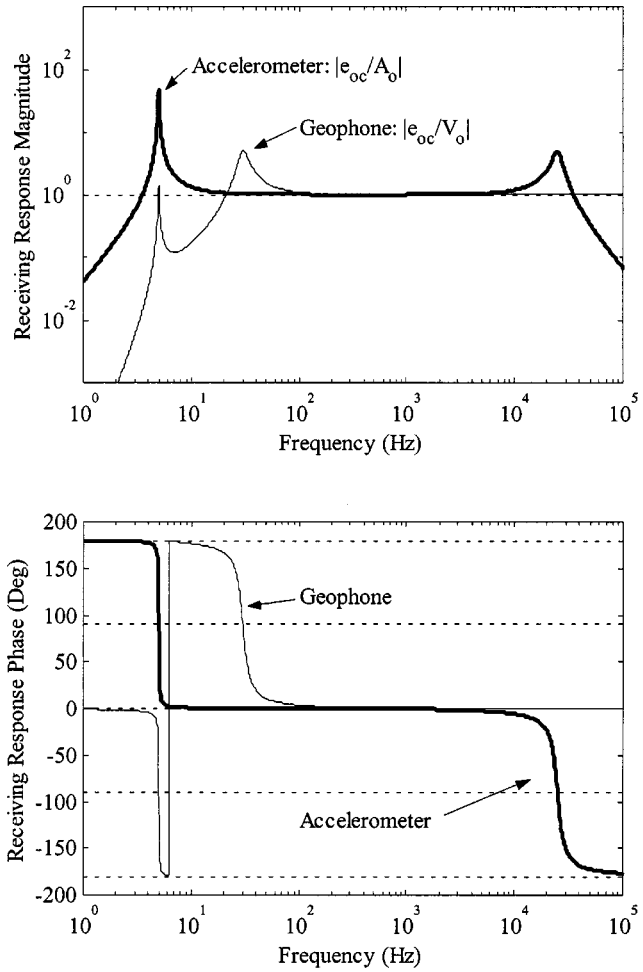


FIG. 8. Open-circuit receiving response of compliantly suspended spherical velocity sensors containing either a moving coil geophone or a piezoelectric accelerometer. The spheres are neutrally buoyant and viscous effects are ignored.

transducer will invariably be much less than the mass of the structure that houses it). The results are plotted in Fig. 8 and normalized so that the amplitude response is unity in the passband. In this way, assumptions do not have to be made regarding the scalar values for ϕ, C_o, and *Bl*. It can be concluded from Fig. 8 that, in the passband, the sensors respond as if they were unconstrained and have simplified receiving responses of $e_{oc}/A_o=\gamma M_a$ and $e_{oc}/V_o=\gamma M_v$, where M_a and M_v are the intrinsic voltage sensitivities of the accelerometer and geophone, respectively, and $\gamma=3\rho_o/(2\rho_s+\rho_o)$ from Eq. (1). In practice, the parameter γ serves as a correction factor to adjust the receiving response of the sphere to a value that is higher, lower, or unchanged from that of the inertial transducer that is imbedded inside the sphere.

D. Wave effects in sphere and suspension

Figure 8 shows that the passband for either sensor spans four orders of magnitude. This is an idealization of what typically occurs in practice, since wave effects in the sphere and the suspension can manifest themselves in the receiving response as spurious resonances. That is, a bare geophone or accelerometer is typically encased in a small, lightweight structure having a high modulus of elasticity. This precludes the structure itself from creating spurious modes in the measurement band; however, once the transducer is imbedded into an external structure such as a sphere, cylinder, or disk made of syntactic foam, polyurethane, etc., then structural modes can become evident and consequently degrade sensor performance. For example, Gabrielson *et al.*[6] showed that a geophone cast inside a cylinder of syntactic foam measuring 3.5 cm in diameter and 8.1 cm long exhibited a resonance close to 2 kHz. Additionally, the (2,1) mode of a hard rubber sphere (e.g., the mode when the sphere deforms into alternating shapes of oblate and prolate spheroids) having a diameter of 5.08 cm, density of 1100 kg/m^3, elastic and shear moduli equal to 2.3 GPa and of 0.1 GPa, respectively, and a Poisson's ratio of 0.4 is approximately 2 kHz when immersed in water. This result is based on the formulas presented by Vogt *et al.*[19] and can be extended to the case when a small inertial transducer is imbedded inside the sphere. For completeness, the transcendental equation which yields the resonance frequencies of a freely vibrating solid elastic sphere is defined as $A_nB_n-C_nD_n=0$,[19] where

$$
\begin{aligned}
A_n&=(k_sa)^2j_n''(k_sa)+(n-1)(n+2)j_n(k_sa),\\
B_n&=(k_ca)^2\left[\left(\frac{\sigma}{1-2\sigma}\right)j_n(k_ca)-j_n''(k_ca)\right],\\
C_n&=(k_ca)j_n'(k_ca)-j_n(k_ca),\\
D_n&=2n(n+1)[j_n(k_sa)-(k_sa)j_n'(k_sa)].
\end{aligned}
\tag{13}
$$

In Eq. (13), k_c and k_s are the compressional and shear wave numbers for the material, j_n is the spherical Bessel function

TABLE I. First two roots of the transcendental equation for a Bernoulli–Euler mass-loaded cantilever beam. The roots are expressed as a function of the mass ratio between the beam mass m and the concentrated mass at the end of the beam M.

m/M	η_1	η_2	$(\eta_2/\eta_1)^2$
0.001	0.23	3.93	292.0
0.01	0.42	3.93	87.6
0.1	0.74	3.94	28.3
1	1.25	4.03	10.4

of the first kind of order n (primes indicate derivatives with respect to the argument), and σ is the Poisson's ratio for the material. The structural wave numbers can be expressed in terms of the acoustic wave number associated with the fluid that surrounds the sphere so that the transcendental equation can be reduced to one independent variable, e.g., $k_c a = (c/c_c)ka$ and $k_s a = (c/c_s)ka$, where c_c and c_s are the compressional and shear wave speeds in the material and $k = \omega/c$ is the acoustic wave number. The structural wave speeds are further defined as $c_c = (E/\rho_s)^{1/2}$ and $c_s = (G/\rho_s)^{1/2}$, where E and G are the elastic and shear moduli of the material.

Wave effects in external suspensions can be analyzed from the standpoint of modeling the system as a mass-loaded Bernoulli–Euler cantilever beam. The beam represents the suspension, and the concentrated mass at the end of the beam represents the sphere and the induced fluid mass. If the internal sensor dynamics are neglected from the analysis, then the resonance frequencies of the system are given by Shabana[20] to be

$$f_n = \frac{1}{2\pi}\frac{\eta_n^2}{L^2}\sqrt{\frac{EI}{\rho A}}, \tag{14}$$

where L, E, I, ρ, and A are the length, elastic modulus, area moment of inertia, density, and cross-sectional area of the beam, respectively. The parameter η_n corresponds to the nth root of the transcendental equation that governs the response of the system, namely[20]

$$k_b L \frac{\cosh k_b L \sin k_b L - \sinh k_b L \cos k_b L}{\cosh k_b L \cos k_b L + 1} = \frac{m}{M}, \tag{15}$$

where k_b is the bending wave number and m/M is the ratio of the beam mass (e.g., $m = \rho AL$) to the concentrated mass (e.g., $M = m_s + m_i$). Note that Ref. 20 has incorrectly expressed Eq. (15) with a minus sign in front of the mass ratio term.

Table I shows the first two roots of Eq. (15) for mass ratios that span three orders of magnitude. The data indicate that the frequency of the fundamental mode varies inversely with the concentrated mass; however, the frequency of the second mode is relatively insensitive to the mass in the system. Wave effects occur at the frequency of the second mode, and recognizing that the modal frequency ratio corresponds to $f_{n+1}/f_n = (\eta_{n+1}/\eta_n)^2$, provides a nondimensional expression that can be used to determine the inception point. For example, if $f_1 = 10$ Hz and $m/M = 0.001$, then $f_2 \approx 3$ kHz.

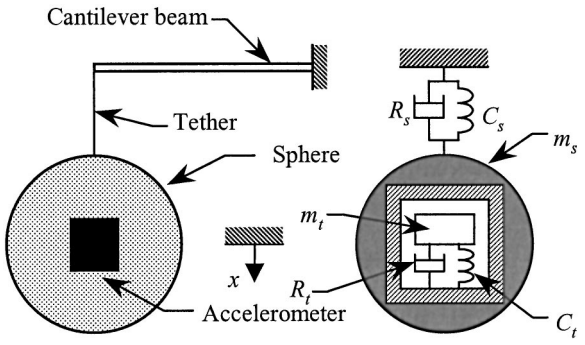


FIG. 9. Basic design of compliantly suspended spherical velocity sensor in its physical and lumped parameter representations.

Now, since most commercially available accelerometers have a flat response from about 10 Hz to 10 kHz, wave effects in the suspension can become a serious design constraint for underwater sensors that rely on the inertial principle to measure the acoustic particle velocity. Damping treatments can reduce the impact from these modes, but should be used with caution since a heavily damped system has the propensity to degrade sensor performance, particularly with respect to the phase response.

III. EXPERIMENTS PERFORMED ON SPHERICAL VELOCITY SENSORS

A. Sensor design and fabrication

Positive, negative, and near-neutrally buoyant spherical sensors are fabricated and tested so that the theory presented in the previous section can be evaluated. More specifically, commercially available birch wood, bronze, and polyurethane spheres, having a diameter of 5.08 cm, are machined to incorporate an Oceana Sensor Technologies model AP1BCN accelerometer.[21] These accelerometers have a nominal sensitivity of 10 mV/g and a mounted resonance frequency of about 25 kHz. They contain a shear-type sensing element and require constant current dc power for an internal FET-based preamplifier circuit. They are ideally suited for this application because they are designed to be imbedded into structures for vibration monitoring.

Each sensor contains a compliant spring that consists of a small, inextensible tether that connects the body of the sphere to a cantilever beam. Figure 9 shows the basic design of the sensor in its physical and lumped parameter representations. Based on the figure, the system can be modeled as a mass-loaded cantilever beam undergoing rigid body translation in the vertical direction. Now, if the internal sensor dynamics of the accelerometer are neglected and the cantilever beam can be modeled as a discrete linear spring, then the lumped parameter resonance frequency of the system is predicted by[22]

$$f_s = \frac{1}{2\pi\sqrt{(m_s + m_i)C_s}} = \frac{1}{2\pi}\sqrt{\frac{3EI}{(m_s + m_i)L^3}}. \tag{16}$$

Note that the basis for neglecting the internal sensor dynamics is that the resonance frequency of the accelerometer is well above the frequency range of interest.

TABLE II. Physical properties for spherical velocity sensors in water.

Sphere	Buoyancy	m_s (g)	ρ_s (g/cm^3)	γ (−)	Spring	C_s (mm/N)	ζ_s[a] (−)	f_s[a] (Hz)	f_s[b] (Hz)
Birch wood	Positive	40.0	0.58	1.39	Steel rod	4.74	0.02	8.8	8.5
Bronze	Negative	499.0	7.27	0.19	Brass strip	1.41	0.01	5.3	5.8
Polyurethane	Near neutral	77.5	1.13	0.92	Steel rod	10.5	0.06	4.0	4.7

[a]Experimental results.
[b]Theoretical results via Eq. (16).

The logarithmic decrement method[23] is used to experimentally determine the resonance frequency and damping factor of each sphere. This is done by submerging the sphere inside a vertically oriented column of water and allowing it to vibrate freely once it is displaced from its equilibrium position. The transient time-series data from the accelerometer embedded within the sphere are measured using an HP 35670A dynamic signal analyzer[24] and downloaded to computer for postprocessing.

During the experiment, the *in situ* dimensions of the cantilever beam are measured with a micrometer. Using these parameters in conjunction with the physical properties of the beam and the sphere enables a theoretical prediction to be made using Eq. (16). The results of the experimental and theoretical analyses are presented in Table II. It can be inferred from the table that the resonance frequencies determined by experiment are in good agreement with those predicted by theory, and that the spheres are lightly damped.

Taking the analysis one step further requires the use of Eqs. (14) and (15). Table III presents the results of this evaluation and shows that the frequencies of the fundamental modes are identical to those predicted by Eq. (16), and except for the bronze sphere, all of the higher-order modes occur at frequencies at nominally 1 kHz, or higher.

B. Sensor performance with respect to changes in sphere density

The second part of the experiment concerns the performance of the sensors with respect to changes in the sphere density. The testing is conducted in a standing wave field consistent with the approach taken by Gabrielson *et al.*,[6] in which the velocity of an inertial sensor was compared to that inferred from a reference hydrophone positioned at the same depth within a waveguide having an acoustic source at one end and a pressure-release surface at the other end. The waveguide used for this testing is made of acrylic and has an inside diameter of 10.16 cm, a length of 100 cm, and wall thickness of 3.18 mm. The walls of the waveguide are compliant; thus, the longitudinal sound speed is substantially reduced from the bulk speed found in open water. The sound speed is determined via experiment[25] to be about 381 m/s. This result compares favorably with theory for ''slow'' waveguides.[26–30] Note that a discussion of the procedure to determine the sound speed is the subject of a separate paper to be submitted in the future.

TABLE III. Continuous system analysis of spherical velocity sensor suspension modes with spheres immersed in water.

Sphere	m/M	f_1 (Hz)	f_2 (Hz)	f_3 (Hz)
Birch wood	0.005	8.5	1185	3841
Bronze	0.017	5.8	397	1285
Polyurethane	0.002	4.7	972	3151

Figure 10 presents a schematic of the experimental setup and shows how the spheres and the reference hydrophone are positioned in the waveguide. A Reson TC-4013 hydrophone,[31] having a measured sensitivity of 28.2 μV/Pa $\pm$6%, is used as the reference standard. The sensitivity is determined prior to testing using a B&K 4229 hydrophone calibrator equipped with a B&K UA0548 coupler.[32] For each test case the measurement depth is 7.62 cm below the air-water (e.g., pressure-release) surface. The sound projector at the bottom of the waveguide is a Wilcoxon Research F4 shaker[33] that is modified to incorporate a 7.62-cm-diameter aluminum plane piston.

Under the assumption of a lossless waveguide having a pressure-release boundary condition, the pressure and particle velocity at some depth d from the free surface are $p(d)=P_o \sin kd$ and $v(d)=j(P_o/\rho_o c)\cos kd$, where P_o is the pressure amplitude and c and k are the in-duct sound speed and longitudinal wave number, respectively. Using these formulas in conjunction with the velocity of the sphere at depth d (e.g., $V_s=A_s/j\omega$) leads to

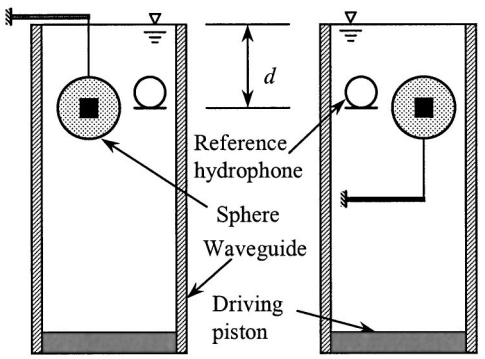


FIG. 10. Experimental setup for tests performed in a fluid-filled waveguide.

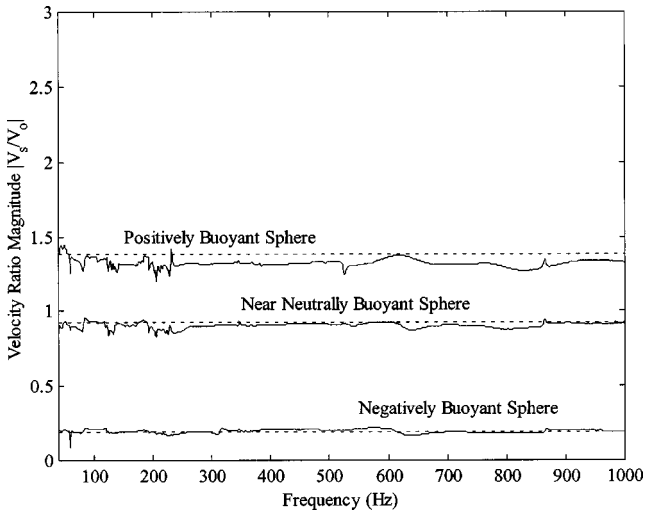


FIG. 11. Magnitude response of spheres tested in a water-filled waveguide versus theory predicted by Eq. (1). The theoretical predictions are plotted as dashed lines.

$$\frac{V_s}{V_o} = -\frac{A_s}{p(d)} \cdot \frac{\rho_o c \tan kd}{\omega}, \tag{17}$$

relating the magnitudes of the sphere velocity and measured acceleration, V_s and A_s, to the local acoustic particle velocity and measured pressure, V_o and $p(d)$. Equation (17) is implemented by measuring the transfer function between the output voltages generated by the accelerometer in the sphere and the reference hydrophone. This is done using an HP 35670A analyzer operating in FFT-analysis mode. The waveguide is ensonified with bandlimited random noise over the 0–1600-Hz frequency range and the transfer function is measured using a 1-Hz analysis resolution, 128 rms averages, and a Hanning window. The transfer function is subsequently adjusted by the sensitivities of the transducers. Note that, prior to the experiment, the vendor-specified receiving sensitivity of each accelerometer is verified to ±3% using a PCB 394C06 handheld shaker.[34]

The results of the experiment are presented in Fig. 11, which compares Eqs. (1) and (17) over the 40–1000-Hz frequency range. This figure shows that the experimental data track well with theory, in that varying the density of the sphere causes a predictable change in the corresponding velocity amplitude. Statistical analysis of the test data reveals that, on average, the response of the spheres deviates from theory by up to 5%. The phase response is omitted from this analysis because this would have necessitated an absolute calibration of the accelerometer and the reference hydrophone.

It is noted that the analysis of the data shown in Fig. 11 is limited to the band of 40–1000 Hz. This is done because: (i) the F4 shaker, in its as-assembled configuration, could not provide sufficient levels of excitation below about 40 Hz; (ii) in Eq. (17), the slope of $\tan kd$ changes rapidly for frequencies greater than 1000 Hz and consequently puts excruciating demands on the accuracy of the measured sound speed, depth, and knowledge of any losses in the waveguide; and (iii) depending on the material, wave effects in the sphere occur at approximately 2 kHz.

It is also noted that the in-band suspension modes shown in Table III do not appear to be evident in the test data. This is believed to be the result of assuming the tether to be inextensible over the entire frequency range of interest. To be more precise, the tether is more aptly modeled as a compliant spring that connects the sphere to the cantilever beam. For the case of the fundamental mode, the system exhibits rigid body motion such that the cantilever beam and end mass move in phase. By extension, it is thought that the compliance of the tether is much less than that of the cantilever beam, and as such, the compliance of the beam dictates the frequency of the fundamental mode. However, for the case of the higher-order modes, where wave effects cause portions of the beam to move at some phase relationship relative to the sphere, the compliance of the tether dominates and essentially decouples the sphere from the beam. In other words, the impedance of the suspension tends to zero (e.g., $R_s + 1/j\omega C_s \rightarrow 0$) and the sphere becomes unconstrained.

C. Sensor performance with respect to changes in fluid density

The experiment described in Sec. II B is repeated to obtain data on the polyurethane sphere as a function of fluid density variation. More specifically, the sphere is subjected to a standing wave field using working fluids other than water. In one case glycerin is used and in another case air is used. The sphere is positively buoyant in glycerin such that $\gamma = 1.07$ and is negatively buoyant in air such that $\gamma = 0.0016$.

Now, since glycerin can hardly be considered an inviscid fluid, Eq. (6) is evaluated at the lowest frequency of interest, e.g., 40 Hz, to assess whether the data should be corrected for viscous effects. The material properties of glycerin at 1 atm and 20 °C are $\rho_o = 1260\ \text{kg/m}^3$ and $\nu = 0.005\ \text{m}^2/\text{s}$. Using these parameters along with those delineated in Table II for the polyurethane sphere leads to $\beta = 0.1$ and $|V_s/V_o| = 1.05$. The latter of these two parameters deviates from Eq. (1) (e.g., $\gamma = 1.07$) by 2%; thus, the effect of viscosity is considered negligible. Figure 12 shows the result of testing the sphere in the glycerin-filled waveguide at a depth $d = 3.81$ cm below the pressure-release surface. The data acquisition and post-test data analysis procedures are identical to those described in the previous section. Figure 12 shows that the experimental data are in very good agreement with theory. Moreover, statistical analysis of the data shows that, on average, the experimental data deviates from theory by about 1%. It is noted that when the waveguide employs glycerin as the working fluid, the measured longitudinal sound speed is about 344 m/s. Consequently, the concerns regarding the $\tan kd$ function in Eq. (17) become more severe relative to the testing performed in water. In order to circumvent any errors, the testing performed in glycerin is done at a slightly shallower measurement depth.

When air is used as the working fluid in the waveguide, the pressure-release boundary condition is invalidated. In this case, the radiation impedance at the open end of the tube must be integrated into equation that relates the sphere velocity to the acoustic particle velocity in the waveguide.

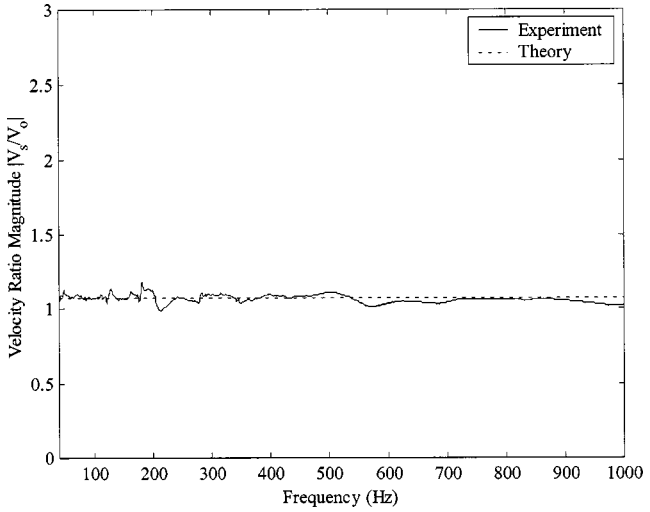


FIG. 12. Magnitude response of polyurethane sphere tested in a glycerin-filled waveguide versus theory predicted by Eq. (1). The theoretical prediction is plotted as a dashed line.

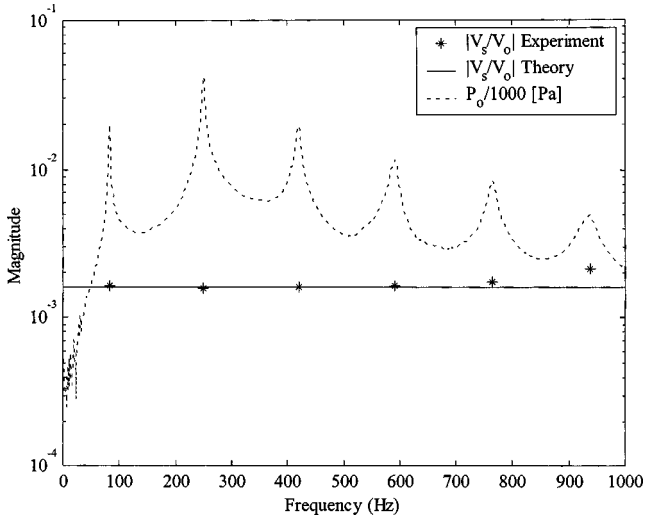


FIG. 13. Magnitude response of polyurethane sphere tested in an air-filled waveguide versus theory predicted by Eq. (1). The calibrated power spectrum of the microphone measurement is also plotted for reference.

More specifically, Kinsler *et al.*[35] relate the particle velocity to the pressure via the specific acoustic impedance, e.g., $V_o = p(d)/Z(d)$, where

$$Z(d) = \rho_o c \frac{1 + Re^{-j2kd}}{1 - Re^{-j2kd}},$$

$$R = \frac{Z_{\text{rad}}/\rho_o cS - 1}{Z_{\text{rad}}/\rho_o cS + 1}, \tag{18}$$

$$\frac{Z_{\text{rad}}}{\rho_o cS} = \frac{(ka)^2}{4} + j0.6ka.$$

In Eq. (18), R is the complex reflection coefficient, Z_{rad} is the mechanical radiation impedance for an open-ended circular waveguide having an unflanged termination, S is the cross-sectional area of the waveguide, and d is the measurement depth relative to the open end. The foregoing leads to $V_s/V_o = A_s Z(d)/j\omega p(d)$.

The testing performed in the air-filled waveguide required the use of a microphone instead of a hydrophone as the reference sensor. Accordingly, a TMS 140AC free-field microphone equipped with a TMS 26AJ preamplifier and a TMS 112AA power supply is used.[36] Prior to the experiment, the sensitivity of the microphone is determined to be 15.36 mV/Pa$\pm$1% using a Quest Technologies QC-20 calibrator.[37]

Frequency response function data between the accelerometer in the sphere and the microphone are obtained with an HP 35670A analyzer operating in swept-sine mode over the 1-Hz to 1000-Hz frequency range. The data are acquired using a 1-Hz analysis resolution and a Hanning window. This is done, as opposed to using bandlimited random noise excitation, to minimize the signal-to-noise error associated with setting an extremely negatively buoyant sphere into motion with a relatively small acoustic driving potential. Figure 13 presents the result of the in-air testing and shows that the data obtained at the waveguide's longitudinal resonance frequencies (e.g., 83, 250, 420, 592, 765, and 937 Hz) compare favorably with theory except at frequencies approaching 1 kHz. The data obtained off-resonance are not presented because the motion of the sphere was masked by the noise floor of the accelerometer. Note that the measurement depth is $d = 7.62$ cm and the longitudinal sound speed is taken to be 343 m/s. These parameters indicate that the sensitivity of the $\tan kd$ function is in part responsible for the lackluster performance at 937 Hz. The calibrated pressure spectrum obtained with the microphone is also presented in Fig. 13 for reference.

It can be concluded from Figs. 12 and 13 that Eq. (1) is also valid for the case of subjecting a sphere to acoustic excitation in fluids having densities that are greater or less than that of the sphere. Also, viscous effects are negligible since the value of $\beta \ll 1$.

IV. CONCLUSIONS

The fidelity of spherical acoustic velocity sensors that rely on the inertial transduction principle is affected by many factors. The fundamental resonances in the system set the passband of the device, and the density of the sphere and the surrounding fluid medium dictate any corrections to the intrinsic voltage sensitivity. Corrections may also need to be applied when viscous effects cannot be ignored. Ideally, the transducer designer tries to develop a sensor that can faithfully reproduce the amplitude and phase response of some acoustic wave of interest; however, the aforementioned constraints can prove this effort to be somewhat challenging for sensors that require a large bandwidth. Wave effects in the sphere and the external suspension present other design issues. Work continues in this area, particularly with respect to developing sensors that have nonspherical geometries (e.g., cylindrical). Additionally, a more in-depth analysis of elastically coated sensors is warranted.

ACKNOWLEDGMENTS

This work was funded by the Naval Air Warfare Center Aircraft Division via the DoD SBIR program (Topic: OSD96-013, Contracts: N68335-97-C-0102 and N00421-98-

C-1234, Technical Monitor: Mr. Gordon K. Marshall). The accelerometers used in the spherical sensors were donated by Oceana Sensor Technologies. Tony Bontomase of Acoustech assisted with the setup of some of the experiments. The author would like to thank Professor Gerald C. Lauchle and Professor Thomas B. Gabrielson of The Pennsylvania State University and Dr. Robert C. Haberman of BBN Technologies for their helpful comments.

[1] E. Skudrzyk, *The Foundations of Acoustics* (Springer, New York, 1971), pp. 361–363.
[2] C. B. Leslie, J. M. Kendall, and J. L. Jones, ''Hydrophone for measuring particle velocity,'' J. Acoust. Soc. Am. **28**, 711–715 (1956).
[3] G. K. Batchelor, *An Introduction to Fluid Dynamics* (Cambridge University Press, New York, 1994), p. 454.
[4] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed. (Butterworth-Heinemann, Oxford, England, 1995), pp. 30–31.
[5] N. A. Fuchs, *The Mechanics of Aerosols* (Dover, New York, 1989), pp. 83–84.
[6] 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).
[7] M. B. Moffett, D. H. Trivett, P. J. Klippel, and P. D. Baird, ''A piezoelectric, flexural-disk, neutrally buoyant, underwater accelerometer,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**(5), 1341–1346 (1998).
[8] 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).
[9] AIP Conference Proceedings 368, *Acoustic Particle Velocity Sensors: Design, Performance, and Applications*, edited by M. J. Berliner and J. F. Lindberg (American Institute of Physics, Woodbury, NY, 1995).
[10] *Proceedings of the Workshop on Directional Acoustic Sensors* (Naval Undersea Warfare Center Division Newport, Newport, RI, 2001).
[11] ANSI S12.12-1992 (R1997), ''Engineering Method for Determining Sound Power Levels of Noise Sources Using Sound Intensity,'' American National Standards Institute, Washington, DC (1997).
[12] ASTM E1050-98, ''Standard Test Method for Impedance and Absorption of Acoustical Materials Using a Tube, Two Microphones, and a Digital Frequency Analysis System,'' American Society for Testing and Materials, Philadelphia, PA (1998).
[13] F. J. Fahy, *Sound Intensity*, 2nd ed. (E & FN Spon, London, 1995), pp. 143–148.
[14] R. S. Woollett, ''Diffraction constants for pressure gradient transducers,'' J. Acoust. Soc. Am. **72**(4), 1105–1113 (1982).
[15] R. J. Bobber, *Underwater Electroacoustic Measurements* (Peninsula, Los Altos, California, 1988), pp. 273–283.
[16] C. L. Morfey, *Dictionary of Acoustics* (Academic, San Diego, 2001), pp. 196 and 407.
[17] R. D. Blevins, *Formulas for Natural Frequency and Mode Shape* (Krieger, Malabar, FL, 1995), pp. 405–409.
[18] S. Temkin, *Elements of Acoustics* (Acoustical Society of America, New York, 2001), p. 446.
[19] R. H. Vogt and W. G. Neubauer, ''Relationship between acoustic reflection and vibrational modes of elastic spheres,'' J. Acoust. Soc. Am. **60**(1), 15–22 (1976).
[20] A. A. Shabana, *Vibration of Discrete and Continuous Systems*, 2nd ed. (Springer, New York, 1997), p. 218.
[21] Oceana Sensor Technologies, Inc., 1632 Corporate Landing Parkway, Virginia Beach, VA 23454.
[22] W. T. Thompson, *Theory of Vibration with Applications*, 3rd ed. (Prentice Hall, Englewood Cliffs, NJ, 1988), p. 20.
[23] R. K. Vierck, *Vibration Analysis*, 2nd ed. (Harper & Row, New York, 1979), pp. 74–76.
[24] Hewlett-Packard Company, Measurement Solutions Division, 8600 Soper Hill Road, Everett, WA 98205.
[25] J. A. McConnell, K. J. Bastyr, and G. C. Lauchle, ''Sound-speed determination in a fluid-filled elastic waveguide,'' J. Acoust. Soc. Am. **105**, 1143–1144(A) (1999).
[26] M. C. Junger and D. Feit, *Sound, Structures, and Their Interaction* (Acoustical Society of America through the American Institute of Physics, Woodbury, NY, 1993), pp. 37–40.
[27] J. Lighthill, *Waves in Fluids* (Cambridge University Press, New York, 1996), pp. 89–100.
[28] J. Parmakian, *Waterhammer Analysis* (Dover, New York, 1963), pp. 8–18.
[29] T. C. Lin and G. W. Morgan, ''Wave propagation through fluid contained in a cylindrical, elastic shell,'' J. Acoust. Soc. Am. **28**, 1165–1176 (1956).
[30] R. Skalak, ''An extension of the theory of water hammer,'' Trans. ASME **78**, 105–116 (1956).
[31] Reson, Inc., 300 Lopez Road, Goleta, CA 93117.
[32] Bruel & Kjaer, 2815 Colonnades Court, Norcross, GA 30071.
[33] Wilcoxon Research, Inc., 21 Firstfield Road, Gaithersburg, MD 20878.
[34] PCB Piezotronics, 3425 Walden Avenue, Depew, NY 14043.
[35] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982), pp. 200–204.
[36] The Modal Shop, Inc., 1776 Mentor Avenue, Suite 170, Cincinnati, OH 45212.
[37] Quest Technologies, 1060 Corporate Center Drive, Oconomowoc, WI 53066.

Green's function estimation using secondary sources in a shallow water environment

Philippe Roux[a] and Mathias Fink
Laboratoire Ondes et Acoustique, ESPCI, Université Paris VII, 10 rue Vauquelin, 75005 Paris, France

(Received 29 October 2001; revised 15 March 2002; accepted 23 October 2002)

This work provides a new way to measure the Green's function between two points in an acoustic channel without emitting a pulse by any of the two points. The Green's function between *A* and *B* is obtained from a set of secondary sources in the guide by averaging either the correlation or the convolution of the signals received in *A* and *B*. A theoretical approach based on mode propagation in a monochromatic regime is presented. Results are then extended to the time domain. Estimation of the Green's function is performed numerically in a range-independent and a range-dependent environment. Application to discreet acoustic communications is discussed. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1542645]

PACS numbers: 43.30.Zk, 43.60.Gk [DLB]

I. INTRODUCTION

In a shallow water environment, acoustic communication is limited by reverberation and multipath propagation that generate multiple echoes which interfere with each other. In general, the presence of multiple echoes blurs the information to be transmitted. One way to overcome this difficulty is to work with an array of transducers and to use the impulse response (or Green's function) between every element of the array and the receiver. Experimentally, this is naturally done during a time-reversal experiment: a first stage consists in recording the impulse response between every transducer of a time-reversal mirror (TRM) and the desired focal point. In a second stage, the time-reversal version of the Green's functions acts as a natural filter for acoustic communication without any distortion due to waveguide interfaces. Several authors have studied the efficiency and the robustness of time-reversal focusing[1–4] and its applications to acoustic communication in underwater acoustics.[5,6] The advantage of time-reversal is that it requires no *a priori* knowledge of the propagation medium. Actually, this means that the first stage in the time-reversal process is a learning stage during which the spatial temporal filter between the sources and the receiver is acquired. However, the recording of the impulse response between the time-reversal mirror and the focal point may be a difficult or even an undesirable step in the case of discrete acoustic communication.

In this paper, we present a way to obtain the Green's function between two points in a waveguide without either point emitting a pulse. The Green's function between *A* and *B* is obtained from a set of secondary sources *S* in the guide by averaging either the correlation or the convolution of the signals received in *A* and *B* depending on the relative positions of *A* and *B* with regard to *S*. Because no field is sent by *A* or *B*, this technique would lead to applications in discrete acoustic communication for which source detection is undesirable. Furthermore, this method does not require any knowledge of the waveguide characteristics and can be applied in a range-dependent environment.

The paper is structured as follows: in the first part, a theoretical approach shows a way to obtain the Green's function between two points using a set of secondary sources in the waveguide. The theory is first presented in the monochromatic regime and is then extended to the time domain. In the second part, numerical results are presented in two configurations: a range-dependent and a range-independent oceanic waveguide. The accuracy of the estimated Green's function is discussed in terms of time-reversal focusing and a quantitative comparison with the exact Green's function of the waveguide is performed. In the last section, we describe the practical limitations of this technique in an oceanic environment and we discuss possible applications to discrete acoustic communications in shallow water.

II. THEORETICAL APPROACH

Figure 1 is a schematic of the waveguide from which the theoretical approach is performed. The question we try to answer is: how does one get the Green's function between points *A* and *B* without using *A* or *B* as a source? The sources are a set of vertical sources *S* located between *A* and *B* [Fig. 1(a)] or outside the channel delimited by the points *A* and *B* [Fig. 1(b)]. The sources *S* and the receivers *A* and *B* are coplanar. In the two configurations, the sources *S* cover the whole water column. Practically speaking, *S* could be a unique source sent from a ship or a plane and that slowly drops from the surface to the ocean bottom regularly emitting an acoustic beacon during its descent. A similar approach has been introduced by several authors working on source localization in shallow water environments using a virtual receiver.[7,8] The main idea of these works is to use a reference field to correct for oceanic variability. For source localization, use of a reference source can reduce environmental mismatch problem and effectively localize targets.[9,10] Applied to the configuration described in Fig. 1, it means for example use the field emitted from *A* and recorded on the receivers *S* in order to localize *B*. In this paper, we proceed in

[a]Author to whom correspondence should be addressed; present address: Marine Physical Laboratory, Scripps Institution of Oceanography, UCSD, San Diego; electronic mail: philippe.roux@espci.fr

 0001-4966/2003/113(3)/1406/11/$19.00

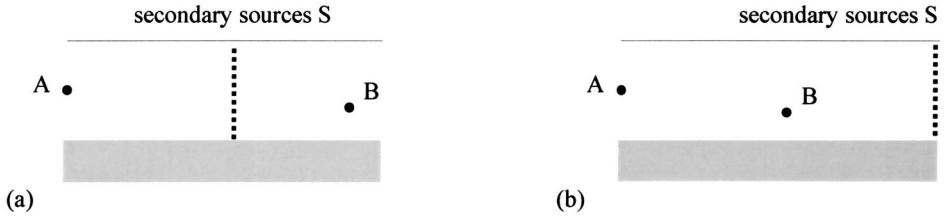


FIG. 1. Schematic of an acoustic waveguide. *A* and *B* are receivers. The secondary sources *S* are distributed over the whole water column: (a) the sources *S* are located between *A* and *B*; (b) the sources *S* are located outside the channel delimited by *A* and *B*.

a reciprocal way: by using an ensemble of secondary sources *S*, we intend to estimate the Green's function between *A* and *B* in the time domain.

For a matter of simplicity, the theoretical approach is described in a monochromatic regime. An acoustic beacon sent from *S* is received by *A* and *B*. In a waveguide, the Green's functions between a source in *S* and receivers in *A* and *B* at a frequency ω are written as follows;

$$G(S,A) \propto \sum_n U_n(A) U_n(S) \frac{\exp(ik_n R_{SA}) \exp(-\alpha_n R_{SA})}{\sqrt{k_n R_{SA}}}, \tag{1}$$

$$G(S,B) \propto \sum_n U_n(S) U_n(B) \frac{\exp(ik_n R_{SB}) \exp(-\alpha_n R_{SB})}{\sqrt{k_n R_{SB}}}, \tag{2}$$

where n is the mode number, $U_n(x)$ is the amplitude of mode n at depth x, k_n and α_n are the propagating and the attenuation wave numbers associated with mode n, respectively, and R_{xy} is the range between x and y. We suppose that the distance R_{AB} is large enough to neglect the contribution to the Green's function $G(A,B)$ of the modal continuum in the penetrable bottom. We make also the approximation that the medium is reciprocal, meaning that fluctuations of the signal due to currents or internal waves are negligible.

A. Configuration 1: The sources *S* are located between *A* and *B*

Multiplying the Green's functions $G(A,S)$ and $G(S,B)$, we get

$$G(A,S)G(S,B) \propto \sum_m \sum_n U_m(A) U_m(S) U_n(S) U_n(B) \times \frac{\exp(ik_m R_{AS}) \exp(-\alpha_m R_{AS})}{\sqrt{k_m R_{AS}}} \times \frac{\exp(ik_n R_{SB}) \exp(-\alpha_n R_{SB})}{\sqrt{k_n R_{SB}}}. \tag{3}$$

We suppose now that the sources *S* are distributed over the whole water column, which allows us to use the orthogonality relation

$$\sum_S U_n(S) U_m(S) \propto \delta_{nm}. \tag{4}$$

In applying orthogonality it has been assumed that the sources *S* span the region where the modes have significant amplitude. In addition density gradients have been neglected. Numerical results to be presented later show that these idealizations have little effect on performance, as long as the sources *S* cover the water column.

Using Eqs. (3) and (4), it follows

$$\sum_S G(A,S)G(S,B) \propto \sum_n U_n(A) U_n(B) \times \frac{\exp(ik_n R_{AB}) \exp(-\alpha_n R_{AB})}{k_n \sqrt{R_{SB} R_{AS}}}, \tag{5}$$

using the relation $R_{AS} + R_{SB} = R_{AB}$.

In Eq. (5), we recognize the amplitude of the modes in *A* and *B* as well as the phase term that describes propagation between *A* and *B*. This leads to the conclusion

$$\sum_S G(A,S)G(S,B) \approx G(A,B) \propto \sum_n U_n(A) U_n(B) \frac{\exp(ik_n R_{AB}) \exp(-\alpha_n R_{AB})}{\sqrt{k_n R_{AB}}}. \tag{6}$$

We note obviously that the spreading factor between Eqs. (5) and (6) is not the same but it is a constant factor that has no importance in the estimation of the Green's function. Actually, the difference between Eq. (5) and the true Green's function $G(A,B)$ appears in the amplitude term for each mode in which k_n should be replaced by $\sqrt{k_n}$. Despite this difference, we will see with numerical examples that $\Sigma_S G(A,S)G(S,B)$ is a good approximation for $G(A,B)$.

B. Configuration 2: The sources *S* are located outside the channel delimited by *A* and *B*

In this case, we multiply the Green's function $G(A,S)$ by the phase-conjugated Green's function $G(S,B)^*$,

$$G(A,S)G(S,B)^* \propto \sum_m \sum_n U_m(A) U_m(S) U_n(S) U_n(B) \times \frac{\exp(ik_m R_{AS}) \exp(-\alpha_m R_{AS})}{\sqrt{k_m R_{AS}}} \times \frac{\exp(-ik_n R_{SB}) \exp(-\alpha_n R_{SB})}{\sqrt{k_n R_{SB}}}. \tag{7}$$

Using Eq. (4) and the relation $R_{AS} - R_{SB} = R_{AB}$, we obtain now

$$\sum_S G(A,S)G(S,B)^*$$

$$\propto \sum_n U_n(A)U_n(B)\frac{\exp(ik_nR_{AB})\exp(-\alpha_nR_{AB})}{k_n\sqrt{R_{SB}R_{AS}}}$$

$$\times\exp(-2\alpha_nR_{SB})\approx G(A,B). \quad (8)$$

As in Eq. (5), we recognize in Eq. (8) the modal amplitude and the phase term of the Green's function between A and B. Compared to $G(A,B)$, we once again note the difference in the spreading factor and the amplitude dependence in $1/k_n$. But the main difference lies now in the amplitude modulation of each mode by an exponential decrease that depends on the mode number through the attenuation wave number α_n. If $\exp(-2\alpha_nR_{SB})$ is roughly constant from mode to mode, this modulation plays no role in the estimation of the Green's function $G(A,B)$ as the spreading factor $1/\sqrt{R_{SB}R_{AS}}$. However, this is generally not the case and α_n classically increases as a function of the mode number. The contribution of the higher-order modes to the estimated Green's function in this configuration will then be less accurate.

III. TIME-DOMAIN SOLUTIONS

In the first configuration, the result obtained in Eq. (6) can be written in the time domain as follows

$$G(A,B,t)\approx\sum_S G(A,S,t)^*G(S,B,t), \quad (9)$$

where the asterisk refers to convolution. Equation (9) has a well-known physical interpretation : it is the application of Huyghen's theorem. The Green's function between A and B can be seen as the sum over a set of points S, that cover a section of the guide, of the Green's function between A and S convolved with the Green's function between S and B.

In practice though, the sources are in S and receivers are in A and B. Invoking reciprocity, we have $G(A,S,t)=G(S,A,t)$. This leads to another formulation equivalent to Eq. (9)

$$G(A,B,t)\approx\sum_S G(S,A,t)^*G(S,B,t). \quad (10)$$

In the second configuration and with the help of reciprocity, the result obtained in Eq. (8) can be written in the time domain as follows:

$$G(A,B,t)\approx\sum_S G(S,A,t)\otimes G(S,B,t), \quad (11)$$

where $\otimes$ refers now to correlation.

Equations (10) and (11) mean that the Green's function between two receivers in a shallow water waveguide can be estimated from a set of secondary sources by either the correlation or convolution of the signals received in A and B.

Equation (11) has recently been introduced in a remarkable paper[11] by Weaver *et al.* They experimentally demonstrated for elastic waves in an aluminum sample that the Green's function between points A and B is obtained from a set of secondary sources S by averaging the correlation of the field, respectively, measured in A and B. What is fascinating in Weaver's approach is that the secondary sources are either acoustic noise from thermal fluctuations inside the sample or deterministic signals from a piezoelectric transducer outside the sample. In underwater acoustics as well as in ultrasonics, this theorem holds only if Eq. (4) is valid. This means, using the language of Weaver, that the total energy sent from the secondary sources S is equidistributed over all the modes that are present in the medium. In other words, every mode must be equally excited after summing over the sources S. We wish also to make a strong connection between this work and a paper from Draeger *et al.*[12] on time reversal in chaotic cavities.

We will see in the following that the Green's function $G(A,B,t)$ can be obtained if the sources S fulfill the following conditions:

(1) the sources S span the whole water column,
(2) the depth between two neighboring sources S is such that the highest-order mode that significantly contributes to the Green's function $G(A,B)$ is correctly sampled. The weight of each mode is related to the exponential dependence of the mode amplitude versus range $\exp(-\alpha_nR_{AB})$ as shown in Eq. (6). The attenuation wave number α_n depends mostly on the attenuation parameter in the bottom.

However, we will show that an estimation of the Green's function between A and B is still possible even if the secondary sources S do not satisfy these conditions.

IV. NUMERICAL RESULTS

The results presented in this paper are obtained with a time-domain simulation performed with a parabolic equation code derived from Michael Collins's RAM code.[13] The emitted signal is a Gaussian-modulated pulse centered at 200 Hz and whose spectrum ranges from 100 to 300 Hz. The field is synthesized from 200 frequencies, which allow for enough zero padding to avoid aliasing problems and at the same time allow for a sufficient number of energy-carrying frequencies to resolve time-domain effects. For the computation, five Padé terms are used. The range step is the smallest wavelength $\lambda_{\min}=5$ m sent from the source and the depth step is $\lambda_{\min}/5=1$ m.

Two shallow water environments are studied: a range-independent and a range-dependent waveguide (Fig. 2). The goal of this work is to determine the time-domain impulse response between points A and B from a set of sources S that are located between or outside the two points. One approach to quantify the quality of the estimated Green's function is then to perform a time-reversal experiment between A and B.

Practically speaking, the first step of our work is to obtain the estimated Green's function $\tilde{G}(A_i,B,t)$ between a receiver in B and a set of 25 receivers A_i using a set of sources S_j and Eq. (10) or (11) (configuration 1 or 2). In order to compare the estimated Green's function $\tilde{G}(A_i,B,t)$ to the true Green's function $G(A_i,B,t)$, we subsequently

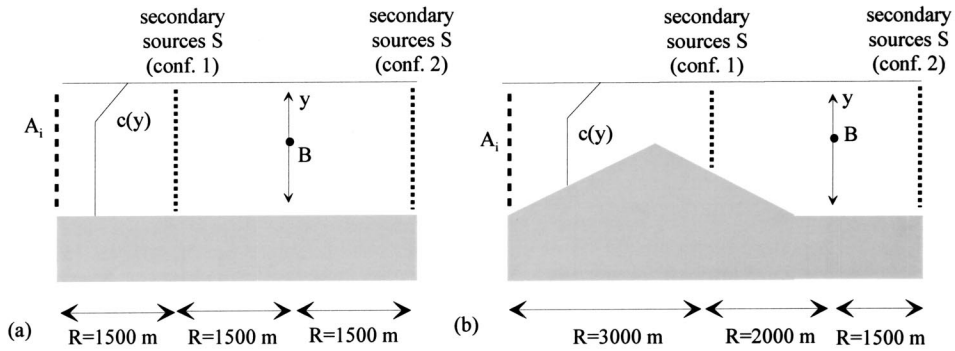


FIG. 2. Description of the acoustic channels used for computation. The sound velocity profile $c(y)$ linearly decreases from 1500 m/s at the surface to 1460 m/s at 25 m and then remains constant in the water column. The bottom density, sound speed, and attenuation are, respectively, 1850 kg/m^3, 1800 m/s, and 0.25 dB/λ; on the left-hand side of the channel, 25 receivers (A_i's) span the water column; (a) range-independent case: water depth=100 m; receiver depth (B) is 40 m; (b) range-dependent case: water depth varies from 100 m on the left-hand side to 50 m at a 2500 m range and then back from 50 to 100 m on the right-hand side; receiver depth (B) is 40 m. Configuration 1 or 2 refers to the positions of the secondary sources S with respect to A and B. Note that for the range-dependent case (b), configuration 2 implies more sources S than configuration 1.

perform a numerical time-reversal experiment between the time reversal mirror made of the A_i's and B. Thus we also consider B and the A_i's as potential sources in order to compare the accuracy of the algorithm that leads to the estimated Green's functions. Time reversal will be performed either with the exact Green's functions or the estimated Green's functions. In the introduction, we described time-reversal as a two-stage process decomposed into a learning process during which the field issued from the source is acquired on the time-reversal mirror, and a backpropagation stage that consists of the transmission of the time-reversed field. Here, the comparison is performed by, respectively, using $\tilde{G}(A_i,B,t)$ or $G(A_i,B,t)$ in the learning stage. It is justified to use time reversal to quantify the accuracy of the estimated Green's functions because it naturally works as a time-domain correlator or, in other words, as a coherent Bartlett beamformer in the time domain. As a consequence, time reversal performed with the estimated Green's functions leads to an ambiguity surface in time and space in which mismatch (leading to high sidelobes) is due to an incorrect estimation of the Green's function. Another heuristic argument in favor of time reversal is the following: in the frequency domain, the Green's function is a complex number with an amplitude and a phase. The correct estimation of the Green's functions requires both amplitude and phase. However they do not have the same relative importance as far as detection, localization, or acoustic transmission is concerned. Thus, an estimated Green's function with an accurate phase and an incorrect amplitude is still useful whereas the contrary is usually useless. A classical least-square difference between the estimated and the exact Green's function is interesting (see Figs. 10 and 11) but does not emphasize the respective roles played by the phase and amplitude errors. On the contrary, phase conjugation in the frequency domain or time reversal with time-dependent signals are a good alternative: if the time-reversed field is focused at the original source, this means that, for each frequency, the phase of the estimated Green's function is exact. Furthermore the sidelobe amplitudes around the source after time reversal is related to the estimation of the amplitude of the Green's function.

Performing time reversal to check the accuracy of the estimated Green's function means that we compute in the time domain

$$\mathrm{Refoc}(y,t)=\sum_{A_i} G(B,A_i,-t)*G(A_i,y,t), \tag{12}$$

and

$$\overline{\mathrm{Refoc}}(y,t)=\sum_{A_i} \tilde{G}(B,A_i,-t)*G(A_i,y,t), \tag{13}$$

where y refers to depth near point B. Roughly speaking, $\mathrm{Refoc}(y,t)$ is the reference time-reversed field around the focal point obtained with the exact Green's functions. This reference time reversal is now compared to $\overline{\mathrm{Refoc}}(y,t)$ which is computed from the estimated Green's functions.

V. RANGE-INDEPENDENT ACOUSTIC CHANNEL

Figure 3 is a spatial temporal representation of the exact Green's functions between B and the A_i's after propagation through the channel. As expected, the pressure field is very dispersed in time due to multiple reflections on the guide interfaces. After propagation through the guide, the field duration is on the order of 0.45 s compared to the 0.02 s duration of the emitted signal. Figure 4 is a spatial temporal

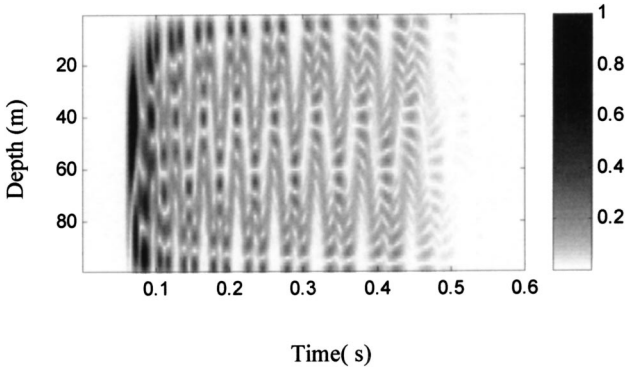


FIG. 3. Spatial–temporal representation of the field received on the TRM made of the A's after transmission from B of a 0.02 s duration pulse. The acoustic channel is described in Fig. 2(a). The source depth is 40 m. Time is along the x axis, the y axis corresponds to depth.

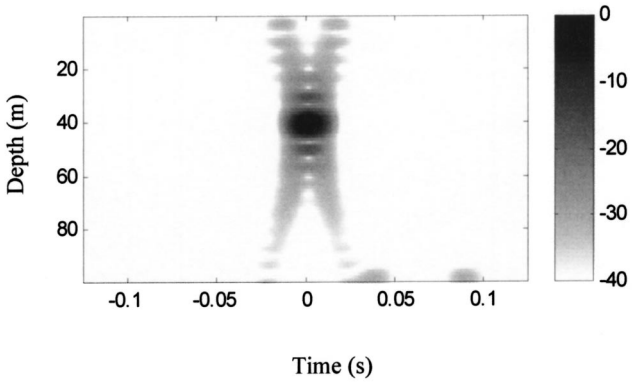


FIG. 4. Spatial–temporal representation in dB of the computed time-reversed field Refoc(y,t) in the plane of receiver B after backpropagation through the channel. The TRM is made of the 25 A_i's. This time-reversed field has been computed with the exact Green's functions following Eq. (12) and is defined as the reference time-reversed field. Time is along the x axis, the y axis corresponds to depth.

representation of the focused field obtained at B after time reversal from the A_i's. As expected in a classical time-reversal experiment, the field is strongly recompressed in time and space at the initial source. Figure 4 is defined as the reference time reversal because it has been computed with the exact Green's functions.

In the following, we first work in configuration 1, i.e., with the secondary sources S placed between A and B [Fig. 2(a)]. In Fig. 5, we present time-reversed fields obtained with the estimated Green's functions. Two points are important. First, the field is still focused in B and the size of the focal spot remains the same whatever the number of secondary sources S used to estimate the Green's functions. This means that a portion of the Green's functions has been estimated with sufficient accuracy over a number of modes to get a narrow focal spot in B. Second, we observe that the sidelobe level is strongly dependent on the number of sources S used to compute the estimated Green's functions. It appears that spatial and temporal sidelobes remain low as long as the depth step between the sources S is equal to or smaller than the smallest wavelength of the source spectrum $\lambda_{\min}=5$ m. We see in Fig. 6 that it is not necessary to sample the water column with too many sources S. The time-reversed field in B presents equivalent temporal sidelobes with sources S distributed with a 2 m depth step or with a $\lambda_{\min}=5$-m depth step. The estimated Green's functions are close to the exact Green's functions in the two cases. The depth step between two neighboring sources S has to be small enough to provide a correct sampling of the highest-order mode that contributes significantly to the Green's function $G(A,B)$. This depends of course on the frequency, on the bottom absorption, and on the range between A and B.

On the other hand, the focal spot is clearly distorted if the sources S do not span the entire water column as shown in Fig. 7. In this case, the Green's functions have not been estimated correctly. We observe that sidelobes are weaker when the sources S are placed near the bottom than near the surface. This is due to the down-refracting profile of the sound speed that has been used for the computation of the pressure field. In a down-refracting profile, refraction bends

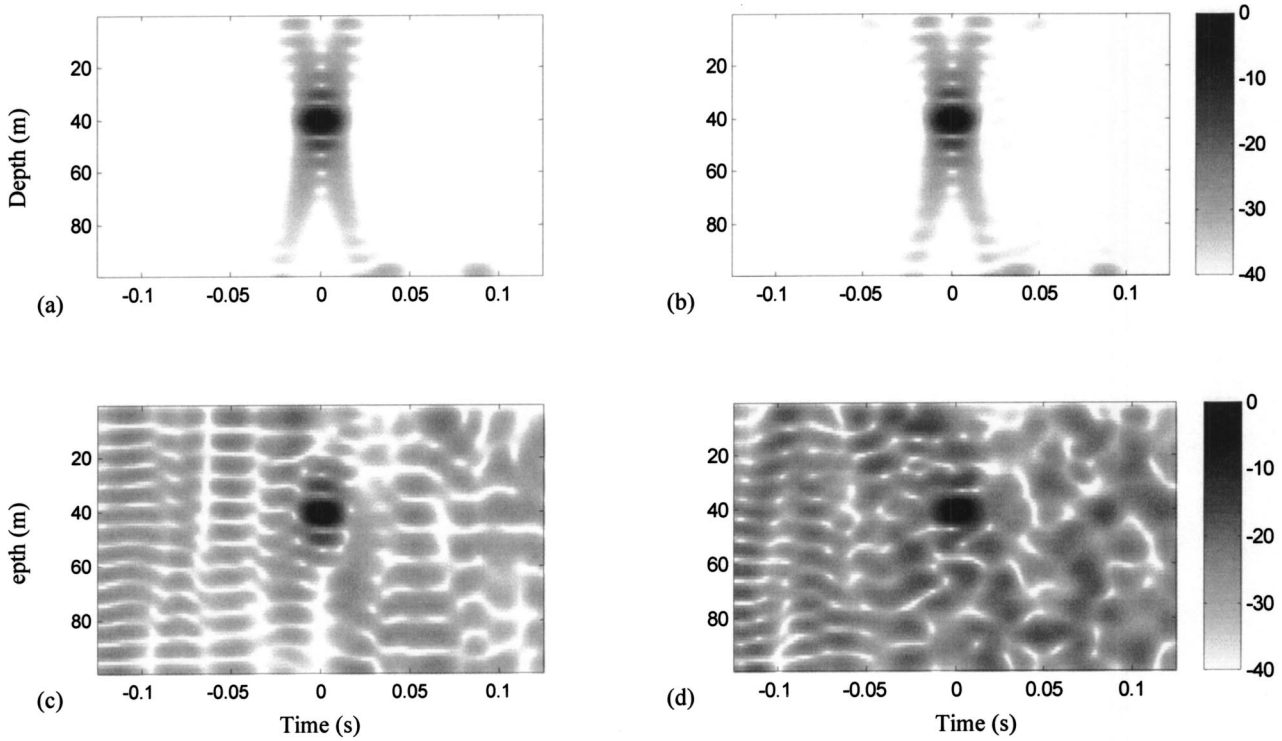


FIG. 5. Spatial–temporal representation in dB of the computed time-reversed field $\overline{\text{Refoc}}(y,t)$ in the plane of receiver B after backpropagation through the channel. The TRM is made of the 25 A_i's. The time-reversed field has been computed with the estimated Green's functions following Eq. (13) and is to be compared to the reference time-reversed field in Fig. 4. The estimated Green's functions has been obtained from Eq. (10) (configuration 1) with a different number of sources S: (a) 50 sources S with a 2 m depth step; (b) 20 sources S with a 5 m depth step; (c) 10 sources S with a 10 m depth step; (d) 5 sources S with a 20 m depth step. Time is along the x axis, the y axis corresponds to depth.

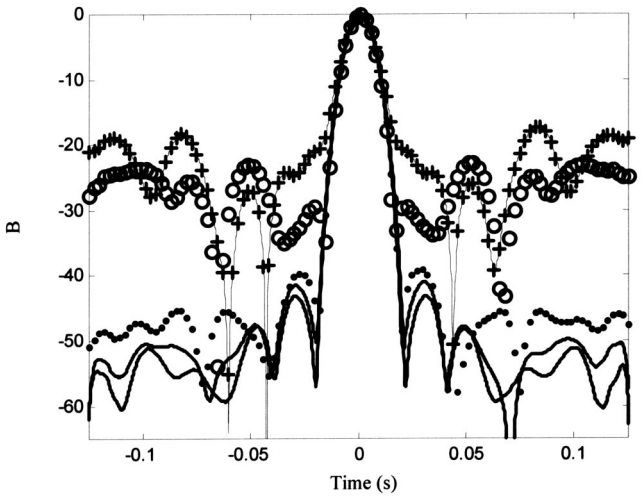


FIG. 6. Comparison in dB between the envelope of the time-reversed fields $\overline{\text{Refoc}}(y,t)$ and Refoc(y,t) in B (receiver depth=40 m). The solid line corresponds to the reference time-reversed field. The estimated Green's functions have been computed with, respectively, 50 sources S with a 2 m depth step (dashed line), 20 sources S with a 5 m depth step (dotted line), 10 sources S with a 10 m depth step (○) and 5 sources S with a 20 m depth step (-+- line). The sources S are placed between A and B (configuration 1).

acoustic rays toward the ocean bottom. As a consequence, the pressure field is better sampled with an array placed near the bottom than near the ocean surface.

According to our theoretical approach, one must satisfy Eq. (4) in order to achieve a good estimation of the Green's functions between the A_i's and B,

$$\sum_S U_n(S)U_m(S) \propto \delta_{nm}. \tag{4}$$

This equation depends only on the number and position of the sources S in the water column. Equation (4) means that, on average, every mode is similarly excited by the sources S in the ocean channel. The average is performed here by the sum over the sources S in Eqs. (10) and (11). A way to confirm this is to check the validity of Eq. (4) in the four configurations presented in Fig. 5. The sources S span the whole water column and the intersource distance is varying from 2 to 20 m. In Fig. 8, we represent the matrix $C_{nm}(S) = \Sigma_S U_n(S)U_m(S)$,[3] for n and m varying from 1 to the total number of modes in the waveguide. The modes U_n have been computed at the central frequency of the source (200 Hz) in the acoustic channel described in Fig. 2(a). We observe that the matrix $C_{nm}(S)$ is close to the identity matrix when the spacing between two sources S is equal or less than the acoustic wavelength. On the other hand, a large depth step between neighboring S sources leads to high off-diagonal elements in the matrix $C_{nm}(S)$ [Figs. 8(c) and (d)] leading to a poor estimation of the estimated Green's functions as shown in Figs. 5(c) and (d).

In the following, we try to quantify results shown in Figs. 5 and 8. To this goal we first measure the difference in terms of least-squares distance between the matrix $C_{nm}(S)$ and the identity matrix for a large set of sources-S that always span the sources S span the whole water column. Figure 9 shows the standard deviation between $C_{nm}(S)$ and matrix identity for a number of sources S varying from 1 to 100. We observe that the standard deviation reaches a floor when the number of sources S is such that the depth step between two sources is on the order of the acoustic wavelength. The floor is nonzero because a portion of the energy of each mode is trapped in the sea-bed which is not sampled by the sources S. This means that Eq. (4) is only approximated when modal penetration into the bottom cannot be neglected.

In Fig. 10, we compare the exact and the estimated Green's functions in the waveguide in the two configurations shown in Fig. 2(a): the sources S are either between points A and B or outside the channel delimited by these two points. When the sources S are between A and B, the Green's function is estimated using Eq. (10) whereas Eq. (11) is used in the other case. For a set of 20 sources S with a 5 m depth step, the two estimated Green's functions are very similar to the exact Green's function. We then extend this comparison by measuring the difference between the exact and the two estimated Green's functions with respect to the number and depths of the sources S. The depths of the sources S are the same as in Fig. 9. We present in Fig. 11 the average standard deviation Std(S) between $\tilde{G}(A_i,B,t)$ and $G(A_i,B,t)$,

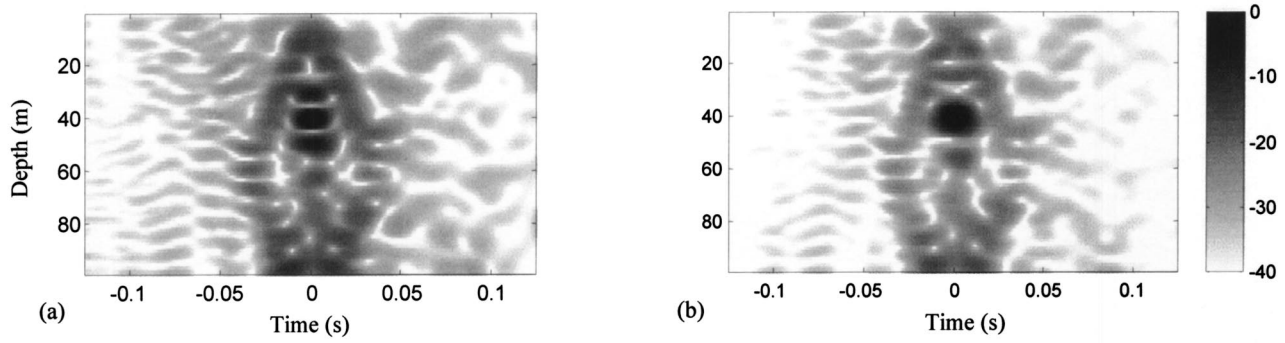


FIG. 7. Spatial–temporal representation in dB of the computed time-reversed field $\overline{\text{Refoc}}(y,t)$ in the plane of receiver B after backpropagation through the channel. The TRM is made of the 25 A_i's. The time-reversed field has been computed with the estimated Green's functions following Eq. (13) and is to be compared to the reference time-reversed field in Fig. 4. The estimated Green's functions have been obtained from Eq. (10) (configuration 1) with 25 sources S that do not span the whole water column: (a) the sources S are distributed between 1 and 49 m with a 2 m step; (b) the sources S are distributed between 50 and 99 m with a 2 m step. Time is along the x axis, the y axis corresponds to depth.

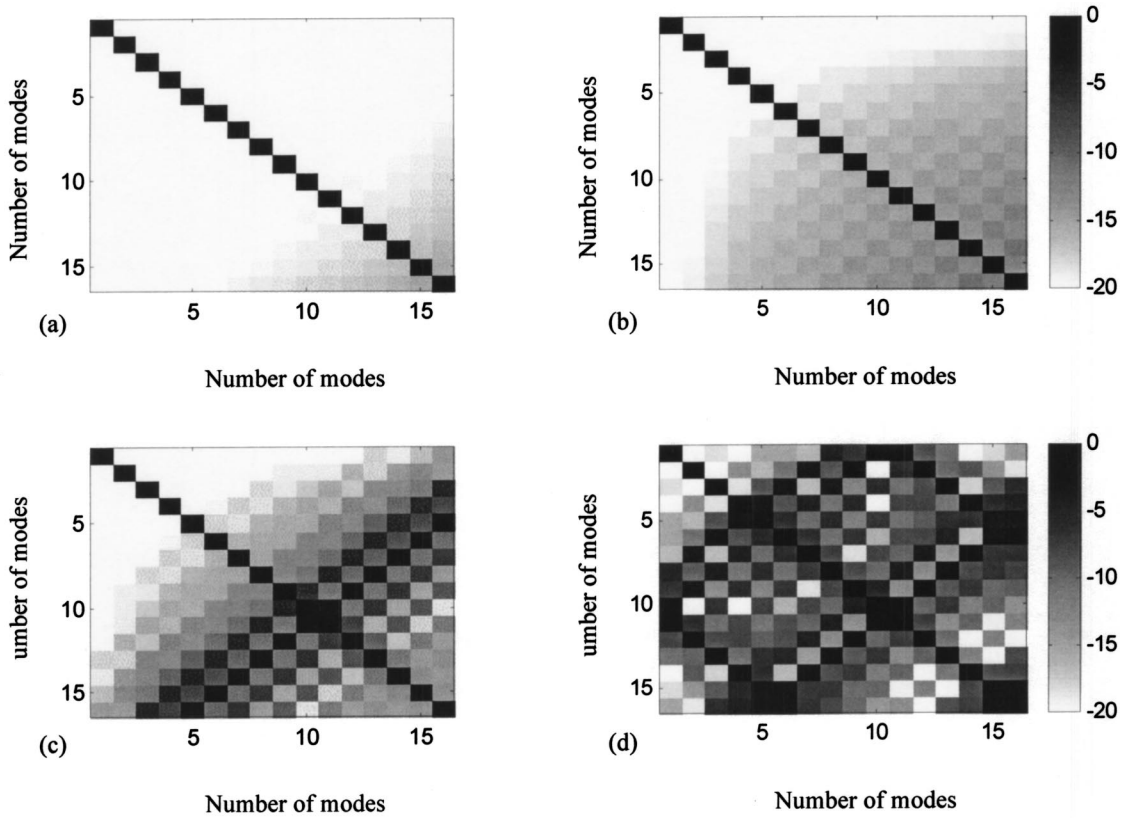


FIG. 8. Comparison in dB between matrices $10\log_{10}(C_{nm}(S))$ computed in different configurations of the sources S: (a) 50 sources S with a 2 m depth step; (b) 20 sources S with a 5 m depth step; (c) 10 sources S with a 10 m depth step; (d) 5 sources S with a 20 m depth step. A good estimation of the Green's function between A and B is obtained when C_{nm} is close to the identity matrix.

$$\mathrm{Std}(S)=\sqrt{\frac{\frac{1}{25}\Sigma_{A_i}\Sigma_t(\tilde{G}(A_i,B,t)-G(A_i,B,t))^2}{\frac{1}{25}\Sigma_{A_i}\Sigma_t G(A_i,B,t)^2}}. \quad (14)$$

Three points are important. First, the results shown in Figs. 9 and 11 are strongly correlated. As expected, the estimated Green's function of Eqs. (10) and (11) are valid if Eq. (4) is satisfied, which provides the respective weight of each mode. In other words, it is possible to get an estimation of the Green's function between two receivers from a set of sources in a waveguide under the condition that every mode is similarly excited by the set of sources. Second, in the two configurations shown in Fig. 2(a), the standard deviation Std(S) reaches a nonzero floor when the number of sources exceeds a certain number (e.g., $N=20$). This means that the estimation of the Green's function is not perfect as noted previously when comparing Eqs. (5)–(8). The amplitude term of the estimated Green's function is at least incorrect by a factor of $\sqrt{k_n}$ for each mode and by an adding factor of exp $(-2\alpha_n R_{SB})$ in configuration 2. However, the phase term of Eqs. (5) and (8) is correct, which means that the estimated Green's function is accurate enough to focus a field using a time-reversal procedure. Indeed, several ultrasonic works have already shown that the amplitude dependence of the field has much less importance than the phase dependence of the field as far as phase conjugation or time reversal is concerned.[14] Last, we see that the floor reached for the standard deviation in Fig. 11 is slightly higher in configuration 2, where the sources S are outside A and B than in configuration 1 where the sources S are between A and B. We show here the influence of the amplitude weight exp$(-2\alpha_n R_{SB})$ that appears for each mode in Eq. (8). The higher modes estimated using Eq. (8) face an extra attenuation that leads to a small extra error in the amplitude of the time domain Green's function.

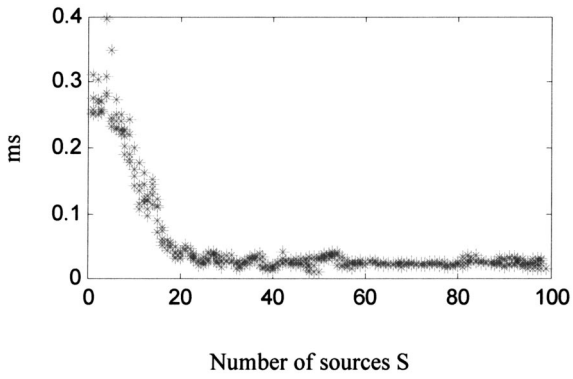


FIG. 9. Standard deviation between $C_{nm}(S)$ and the identity matrix with respect to the number of sources S. For a given number of sources S, several positions of the array of sources S may be possible in the water column leading to one or a few points for each abscissa.

This error would significantly increase for larger distances between S and B, i.e., when the sources S are far away from points A and B.

VI. RANGE-DEPENDENT ACOUSTIC CHANNEL

In the following, we present numerical results obtained in a range-dependent acoustic channel. In a slowly varying

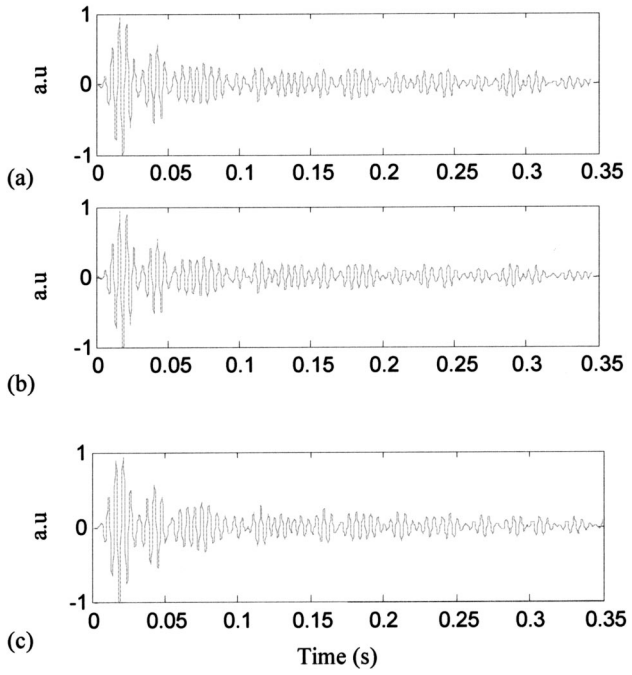


FIG. 10. Range-independent case: comparison between the exact Green's function $G(A,B,t)$ in (a) and the estimated Green's function $\tilde{G}(A,B,t)$ in (b) and (c). (b) and (c) correspond to sources S located between (configuration 1) or outside (configuration 2) A and B, respectively. The depth of points A and B are 37 and 40 m, respectively [cf. Fig. 2(a)]. The estimated Green's functions are computed from 20 sources S with a 5 m depth step.

range-dependent waveguide, Eqs. (1)–(8) can be rewritten using adiabatic mode theory. In this case, the mode shape depends now on the range and each wave number is to be replaced by an average wave number over the propagation range. In this work, we deliberately choose the environment described in Fig. 2(b), which prevents us from applying mode theory or even adiabatic mode theory between A and B. Therefore, Eqs. (1)–(8) do not apply and the Green's function cannot be described using a simple analytic formulation. However, we numerically show that Eqs. (10) and (11) still hold, which means that the Green's functions between B and the A_i's can be estimated from a set of secondary sources S that cover the whole water column.

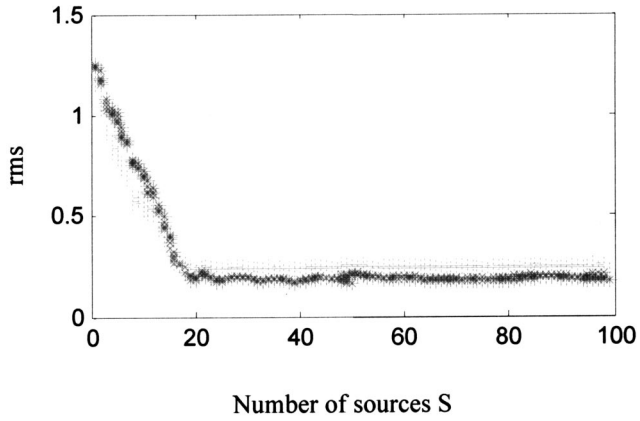


FIG. 11. Range-independent case: standard deviation Std(S) between the estimated and the exact Green's functions on the A_i's with respect to the number of sources S in the waveguide. The dark (*) and the clear (+) correspond to sources S located between (configuration 1) or outside (configuration 2) A and B, respectively. For every set of sources S, Std(S) is averaged over the 25 receivers A_i's. For a given number of sources S, several positions of the array of sources S may be possible in the water column leading to one or a few points for each abscissa.

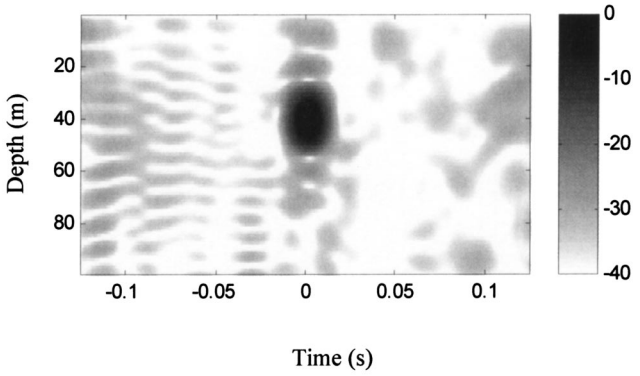


FIG. 12. Spatial–temporal representation in dB of the computed time-reversed field Refoc(y,t) in the plane of receiver B after backpropagation through the range-dependent channel. The TRM is made of the 25 A_i's. This time-reversed field has been computed with the exact Green's functions following Eq. (12) and is defined as the reference time-reversed field. Time is along the x axis, the y axis corresponds to depth.

In Figs. 12 and 13, we present results similar to the results of the range-independent waveguide (Figs. 4 and 5). We first compute a classical time-reversal experiment between B and the A_i's using the exact Green's functions in the time domain. The time-reversed focused field Refoc(y,t) presented in Fig. 12 is once again defined as the reference field. It is interesting to observe that the time-reversed field in B does not present the same pattern as the one obtained in a range-independent waveguide (see Fig. 4). The sidelobe level is higher and the size of the focal spot is twice as large. Actually the higher modes present in A and B no longer exist at the middle of the Chinese-hat shape waveguide and their contribution to the time-reversed focusing has been lost in the sea-bed.

In Fig. 13, we present time-reversed fields obtained with the estimated Green's functions computed from different set of sources S following Eq. (10) (configuration 1). By comparing Refoc(y,t) and $\overline{\text{Refoc}}(y,t)$, we draw the same conclusions as in the case of the range-independent waveguide. Even if we still refocus in B, the estimated Green's functions are close to the exact Green's functions as long as the depth step between two sources S enables a good sampling of the higher-order contributing modes. Of course, the range dependency of the channel prevents correlating the accuracy of the estimated Green's functions to either the modal structure of the waveguide or to the way modes are spatially sampled by the set of sources S.

In Fig. 14, we compare the exact and the estimated Green's functions in the two configurations shown in Fig. 2(b). When the sources S are between A and B, the Green's function is estimated using Eq. (10) whereas Eq. (11) is used in the other case. Keeping the depth sampling constant (5 m step between two neighboring sources S), the number of sources S is not the same in configurations 1 and 2. As in the range-independent case, the estimated and the exact Green's functions present the same temporal dependence.

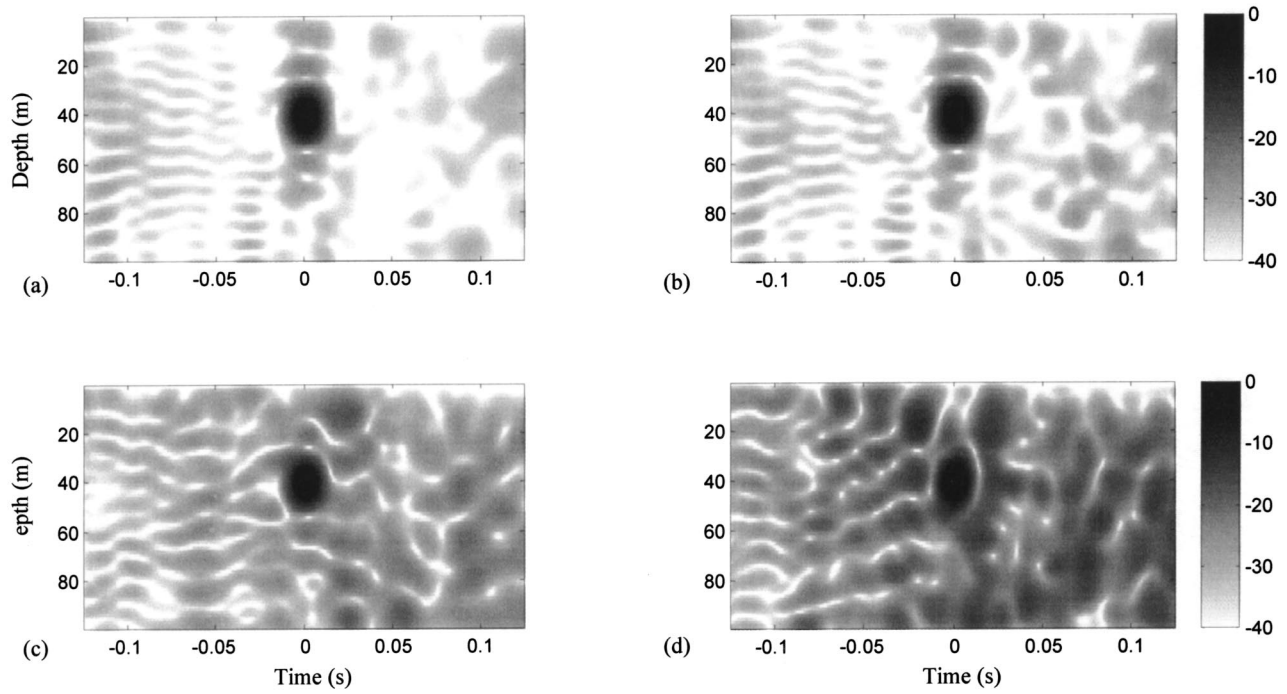


FIG. 13. Spatial–temporal representation in dB of the computed time-reversed field $\overline{\text{Refoc}}(y,t)$ in the plane of receiver B after backpropagation through the range-dependent channel. The TRM is made of the 25 A_i's. The time-reversed field has been computed with the estimated Green's functions following Eq. (13) and is to be compared to the reference time-reversed field in Fig. 12. The estimated Green's functions has been obtained from Eq. (10) (configuration 1) with a different number of sources S: (a) 30 sources S with a 2 m depth step; (b) 12 sources S with a 5 m depth step; (c) 6 sources S with a 10 m depth step; (d) 3 sources S with a 20 m depth step. Time is along the x axis, the y axis corresponds to depth.

We more quantitatively compare the estimated Green's functions to the exact Green's functions by measuring Std(S) in the range-dependent channel. In Fig. 15, we show that results are similar for both the range-dependent and the range-independent waveguides. Two points are important. First, the difference between the estimated Green's functions provided by configuration 1 or 2 relies on the extra attenuation that faces every mode $\exp(-2\alpha_n R_{SB})$ in configuration 2. This error remains small as long as the sources S are not too far away from points A and B. Second, the results obtained with configuration 2 in the range-dependent case shows that Std(S) reaches a floor for 15 sources S whereas 20 sources S

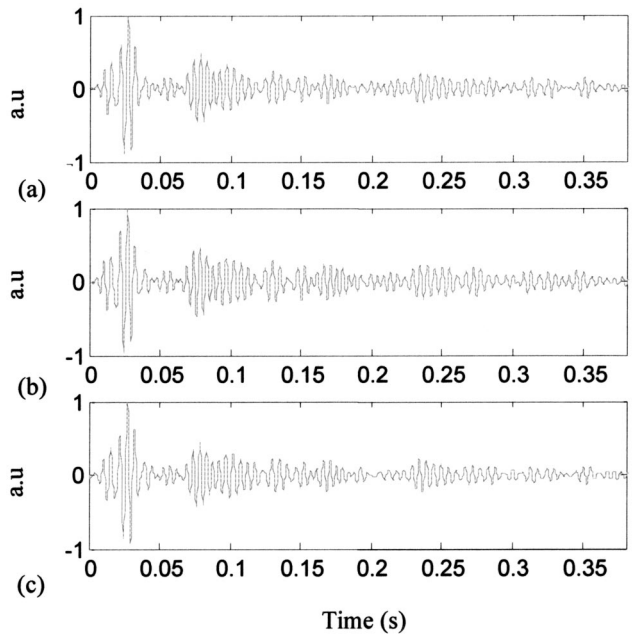


FIG. 14. Range-dependent case: comparison between the exact Green's function $G(A,B,t)$ in (a) and the estimated Green's function $\tilde{G}(A,B,t)$ in (b): (b) and (c) correspond, respectively, to 12 sources S with a 5 m depth step located between A and B (configuration 1) or 20 sources S with a 5 m depth step outside A and B (configuration 2). The depth of points A and B are 37 and 40 m, respectively [cf. Fig. 2(b)].

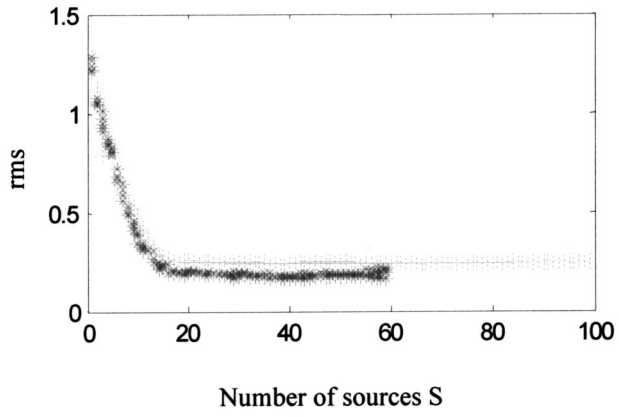


FIG. 15. Range-dependent case: standard deviation Std(S) between the estimated and the exact Green's functions on the A_i's with respect to the number of sources S in the range-dependent channel. The dark (*) and the clear (+) correspond to sources S located between (configuration 1) or outside (configuration 2) A and B, respectively. Note that configuration 2 implies more sources S than configuration 1 [cf. Fig. 2(b)]. For every set of sources S, Std(S) is averaged over the 25 receivers A_i's. For a given number of sources S, several positions of the array of sources S may be possible in the water column leading to either one or a few points for each abscissa.

are necessary in the range-independent case (see Fig. 11). This is due to the fact that the higher-order modes are lost in the Chinese-hat shape bottom present in the range-dependent environment, thus authorizing a less accurate sampling of the water column by the sources *S*.

VII. DISCUSSION

In this work, time-domain numerical simulations using a PE code have been performed in the absence of noise. One may wonder if noise in the ocean could degrade the estimation of the Green's function. The answer is no. Actually, the estimated Green's functions $\tilde{G}(A,B,t)$ are obtained from Eqs. (10) and (11) which imply both either a convolution or a correlation between two fields $G(S,A,t)$ and $G(S,B,t)$ and an average over the set of sources *S*. The correlation, the convolution, and the average are very robust process with respect to random noise. As a consequence, $G(S,A,t)$ and $G(S,B,t)$ may be much more affected by the presence of noise than the estimated Green's function between *A* and *B*.

Of course, the main limitation of this method relies on the location of the secondary sources *S*. We have shown that a good estimation of the Green's function is provided when the sources *S* are distributed along the whole water column. Even in this case, problems arise when the points *A* and *B* are no longer in the same plane as the sources *S*. Indeed, up to now, we have worked in a two-dimensional range-dependent or range-independent waveguide. In this case, sources and receivers belong to the same plane, which implies for example $R_{AS}+R_{SB}=R_{AB}$ for sources *S* placed between the receivers *A* and *B*. However, the ocean is of course a three-dimensional environment and the above-given relation does not hold when the sources *S* and the receivers *A* and *B* do not belong to the same plane. This is a very important condition because the distance $R_{AS}+R_{SB}$ determines the phase of the estimated Green's function as shown in Eqs. (5)–(8). An incorrect estimation of the phase leads to an incorrect estimation of the temporal dependence of the Green's function in the time domain, which means no Green's function at all.

An important concern is the nature and the distribution of the sources *S* in the ocean to get the Green's function between *A* and *B*. It is not easy to get a vertical distribution of sources in the water column. As noted in Sec. I, one possibility could be to use one acoustic source dropped from a plane or a boat and that slowly sinks from the surface to the bottom while emitting repetitively an acoustic beacon during its descent. Furthermore, one could show that a drift of the source during its descent would not degrade the estimation of the Green's function as long as it remains in the *AB* plane. Another artificial source in the ocean is the acoustic noise made by boats. Unfortunately, these sources are distributed horizontally along the boat path and they do not excite similarly all the modes in the waveguide. A last potential acoustic source in the ocean is noise due to surface waves, internal waves, or more generally acoustic noise due to random fluctuations of the ocean. Actually, the first experimental demonstration of the Green's function estimation by secondary sources has been performed using very low thermal fluctuations in an aluminum sample.[11] The advantage of using random fluctuations as a source is the natural equidistribution of the acoustic energy over all the modes of the waveguide. Similarly, in underwater acoustics, it would be interesting to try to estimate the Green's function between points *A* and *B* simply by averaging the correlation of the noise received in *A* and *B*.[15] Experimentally speaking, this only requires broadband receivers with a high sensitivity and a noiseless powerful amplifier. Numerical and experimental works are pursued in this direction by the first author at his new address and by another team in geophysics.[16]

Finally, this work leads to application in discrete underwater acoustic communications. Indeed, beyond the quality of acoustic transmission in shallow water, an important issue is the discretion of the communication between two points in the ocean. As mentioned earlier, it has been shown that time reversal could be a good way to transmit high-rate information through an acoustic channel without being limited by the problem of multipath propagation in such an environment. However, time reversal is a two-step experiment and the first step requires the acquisition of the impulse response between the focal point and the TRM. This stage is not discreet because the source used at the focal point reveals the place to which information is to be transmitted in a second step. On the contrary, the technique described in this paper offers the advantage of using secondary sources to acquire the impulse response between the focal point and the TRM. This allows the focal point to remain discreet in the whole time-reversal experiment. However, acoustic transmission is usually carried out by signals in the kilohertz range, which is far above the spectrum ranges studied in this paper (100–300 Hz). Working at a higher frequency does not change the physical insight described in this work but, depending on the environment, it may practically require a very fine depth-step between two neighboring sources *S* in order to adequately sample the higher-order modes that contribute to the Green's function.

[1] M. Fink, "Time-reversed acoustics," Phys. Today **50**, 34–40 (1997).
[2] M. Fink, "Time-reversed acoustics," Sci. Am. **281**, 67–73 (1999).
[3] W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and D. R. Jackson, "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am. **103**, 25–40 (1998).
[4] H. C. Song, W. A. Kuperman, and W. S. Hodgkiss, "A time-reversal mirror with variable range focusing," J. Acoust. Soc. Am. **103**, 3234–3240 (1998).
[5] P. Roux and M. Fink, "Time-reversal in a waveguide: Study of the spatial and temporal focusing," J. Acoust. Soc. Am. **107**, 2418–2429 (2000).
[6] G. F. Edelmann, T. Akal, W. S. Hodgkiss, S. Kim, W. A. Kuperman, and H. C. Song "An initial demonstration of underwater communication using time-reversal," IEEE Trans. Ultrason. Ferroelectr. Freq. Control (to be published).
[7] M. Siderius, D. R. Jackson, D. Rouseff, and R. Porter, "Multipath compensation in shallow water environments using a virtual receiver," J. Acoust. Soc. Am. **102**, 3439–3449 (1997).
[8] P. D. Mourad, D. Rouseff, R. P. Porter, and A. Al-Kurd, "Source localization using a reference wave to correct for oceanic variability," J. Acoust. Soc. Am. **92**, 1031–1039 (1992).
[9] R. P. Porter, P. D. Mourad, and A. Al-Kurd, "Wave-front reconstruction in variable, multimode waveguides," J. Opt. Soc. Am. A **9**, 1984–1990 (1992).
[10] A. M. Thode, "Source ranging with minimal environmental information using a virtual receiver and waveguide invariant theory," J. Acoust. Soc. Am. **108**, 1582–1594 (2000).

[11] R. L. Weaver and O. I. Lobkis, ''Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies,'' Phys. Rev. Lett. **87**, 134301 (2001).
[12] C. Draeger and M. Fink, ''One-channel time-reversal in chaotic cavities: Theoretical limits,'' J. Acoust. Soc. Am. **105**, 611–617 (1999).
[13] M. D. Collins and E. K. Westwood, ''A higher-order energy-conserving parabolic equation for range-dependent ocean depth, sound speed and density,'' J. Acoust. Soc. Am. **89**, 1068–1075 (1991).
[14] A. Derode, A. Tourin, and M. Fink, ''Ultrasonic pulse compression with one-bit time reversal through multiple scattering,'' J. Appl. Phys. **85**, 6343 (1999).
[15] W. A. Kuperman and F. Ingenito, ''Spatial correlation of surface generated noise in a stratified ocean,'' J. Acoust. Soc. Am. **67**, 1988–1996 (1980).
[16] A. Paul and M. Campillo, ''Extracting the Green's function between two stations from coda waves,'' EOS Trans. Am. Geophys. Union **82–47**, F842 (2001).

Lamb wave reflection at the free edge of a plate

Bruno Morvan,[a] Nicolas Wilkie-Chancellier, Hugues Duflo, Alain Tinel, and Jean Duclos
Laboratoire d'Acoustique Ultrasonore et d'Electronique (L.A.U.E.), CNRS (UMR 6068), Université du Havre, Place Robert Schuman, BP 4006, 76610 Le Havre, France

(Received 6 November 2001; revised 22 November 2002; accepted 22 November 2002)

This study concentrates on the reflection of Lamb waves at the free end of a plate. The conversion phenomena are examined in detail over a large frequency range and the energy conversion coefficients are obtained by three different ways: theoretically, numerically (finite element method) and experimentally. The experimental energy determination is obtained from the measurement of the plate normal displacements, by mean of a laser interferometer. All results are in relatively good agreement and the energy balance between incident and reflected waves is discussed in each case. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1539521]

PACS numbers: 43.35.Cg, 43.35.Zc, 43.40.Dx [DEC]

I. INTRODUCTION

Lamb waves are frequently used in nondestructive testing of plane structures: they allow the detection of surface defects or internal cracks. Because they propagate over a long distance, weakly attenuated (in the case of a free elastic plate), they are used for the evaluation of the assembly of large free plates in aeronautics, for example. Many theoretical or experimental studies deal with Lamb wave propagation in infinite plates. However, manufacturers encounter some difficulties because real plates are limited in length and welded. In this paper, the reflection of Lamb waves at the free edge of a plate is investigated. Torvik[1] has shown that mode conversion phenomena occur to satisfy boundary conditions. A single mode gives rise to several propagating or nonpropagating modes.[1,2] Gregory *et al.*[3] confirm these results in the case of the first symmetric Lamb mode incident at the end of the plate. They develop a ''method of projection'' to compute the repartition of the incident energy between the various reflected waves. Y. Cho *et al.*[4] integrate a boundary element method with the normal mode expansion technique to study the reflection of higher symmetric or antisymmetric incident Lamb waves at the edge. They prove the mode conversions and obtain the reflection coefficients of these waves.

Some experiments identify several mode conversions in a qualitative way.[5] Lowe *et al.* show experimentally that nonpropagating modes occur when the $A0$ mode is reflected from the end of a plate.[6] In this paper, we develop an experimental method to estimate the energy carried out by each reflected wave. The normal surface displacements of the plate are measured with a laser interferometer. Signal analysis allows us to extract accurate displacement values of the incident and reflected Lamb waves from these measurements. These displacements are then processed, so leading to the Lamb wave Poynting vector flow. In order to justify these measurements we develop a theory using the Torvik hypothesis. Moreover, some cases are calculated with the finite element method and verify the main theoretical results. At last, the experimental results are compared with theoretical and finite element method (FEM) ones.

[a]Electronic mail: morvan@iut.univ-lehavre.fr

II. LAMB WAVE PROPAGATING IN A SEMI-INFINITE PLATE

A. Theoretical study of the reflection at the end of a semi-infinite plate

Consider a two-dimensional semi-infinite plate with a free edge on the right side as shown in Fig. 1. This plate is stainless steel (mass density $\rho = 7800\ \mathrm{kg\,m^{-3}}$, velocities $v_T = 3150\ \mathrm{m\,s^{-1}}$ and $v_L = 5840\ \mathrm{m\,s^{-1}}$) and has a thickness d. As Torvik, we make the assumption that the propagating modes in the semi-infinite plate and in an infinite extended plate are identical.[1] An incident harmonic Lamb wave (frequency F) propagating in the $x_1 > 0$ direction is completely reflected at the plate extremity ($x_1 = 0$). If we suppose a single reflected mode, the boundary conditions of zero traction on the edge are not satisfied. This implies that wave conversions and reflected modes are propagating as well as evanescent modes, with real, imaginary or complex wave number, calculated for an infinite plate.[1,2] In other words, at $x_1 = 0$, the incident wave stresses are opposite to the sum of an infinite number of reflected modes stresses:

$$\sum_n r_n T_{11}^n = -T_{11}^{\mathrm{inc}}, \quad \sum_n r_n T_{31}^n = -T_{31}^{\mathrm{inc}}. \tag{1}$$

In these expressions, T_{11}^{inc} and T_{31}^{inc} are the normal and tangential stress components of the incident wave of energy flow ϕ_{inc}; $r_n T_{11}^n$ and $r_n T_{31}^n$ are the respective stresses of the nth reflected wave (if T_{11}^n and T_{31}^n are the stresses associated to a wave of energy φ_n).

In order to obtain the r_n coefficients, the linear system (1) is calculated in Q points regularly set in the thickness d of the plate and for a finite number N of reflected waves:

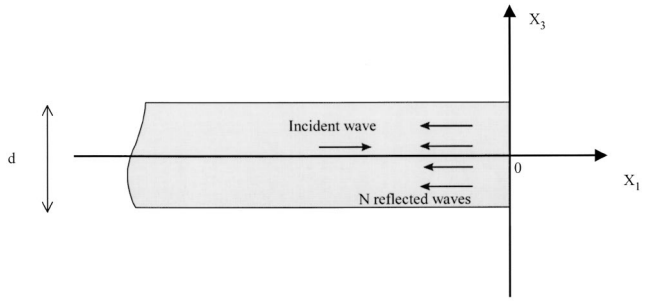


FIG. 1. Lamb wave reflection at the plate extremity.

$$\begin{bmatrix} T_{11}^{\text{inc}}(1) \\ T_{31}^{\text{inc}}(1) \\ T_{11}^{\text{inc}}(2) \\ T_{31}^{\text{inc}}(2) \\ . \\ . \\ . \\ T_{11}^{\text{inc}}(Q) \\ T_{31}^{\text{inc}}(Q) \end{bmatrix} + \begin{bmatrix} T_{11}^{1}(1) & T_{11}^{2}(1) & . & . & . & T_{11}^{N}(1) \\ T_{31}^{1}(1) & T_{31}^{2}(1) & . & . & . & T_{31}^{N}(1) \\ T_{11}^{1}(2) & T_{11}^{2}(2) & . & . & . & T_{11}^{N}(2) \\ T_{31}^{1}(2) & T_{31}^{2}(2) & . & . & . & T_{31}^{N}(2) \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ T_{11}^{1}(Q) & T_{11}^{2}(Q) & . & . & . & T_{11}^{N}(Q) \\ T_{31}^{1}(Q) & T_{31}^{2}(Q) & . & . & . & T_{31}^{N}(Q) \end{bmatrix}$$

$$\times \begin{bmatrix} r_1 \\ r_2 \\ . \\ . \\ . \\ r_N \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ . \\ . \\ . \\ 0 \\ 0 \end{bmatrix}.$$

If we use an evident matrix formulation

$$T + Mr = 0,$$

then

$$r = -(M^T M)^{-1} M^T T,$$

if exponents T and -1 mean respectively transposed and inverse matrix.

The energy carried by the nth reflected wave is written as

$$\phi_n = r_n \cdot r_n^* \varphi_n .$$

In this expression, φ_n is the acoustic Poynting vector flow of the nth Lamb mode. In the particular case of complex or imaginary wave numbers K, the Poynting vector flow φ_n is equal to zero. In fact, these modes exist close to the extremity but do not carry energy along the x_1 direction.

Consequently, the Lamb wave energy reflection coefficient R_n can be calculated by dividing the energy of the reflected wave ϕ_n by the incident energy ϕ_{inc},

$$R_n = \frac{\phi_n}{\phi_{\text{inc}}}.$$

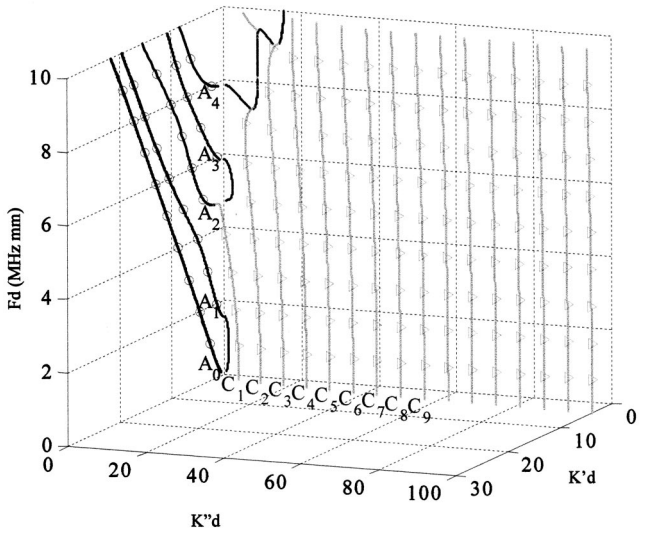


FIG. 2. Real, imaginary and complex antisymmetric wave number roots (○: real; △: complex; — imaginary).

In the next equation, the energy conservation concept can be applied to check the computation accuracy:

$$\phi_{\text{inc}} = \sum_n \phi_n = \sum_n R_n \phi_{\text{inc}} \quad \text{or} \quad \sum_n R_n = 1.$$

The system of equations (1) is computed using 200 points at the plate extremity and 20 reflected modes. These modes are the 20 first real, imaginary and complex modes (at a given frequency, the complex roots are indexed by their increasing imaginary part). We want to test this theory in the case of an $A1$ incident wave. The reflected wave must have the same symmetry so we plot (Fig. 2) the wave number roots $K = K' + jK''$ of the antisymmetric modes. The three dimensional diagram (complex plane) is limited to the positive imaginary and real components of the wave number ($K' > 0$ and $K'' > 0$). In the plane (K', Fd), we recognize the classical Lamb waves dispersion curves. In Fig. 3, the 20 first antisymmetric modes for $Fd = 4$ MHz mm are represented in the complex wave number space. Only the modes

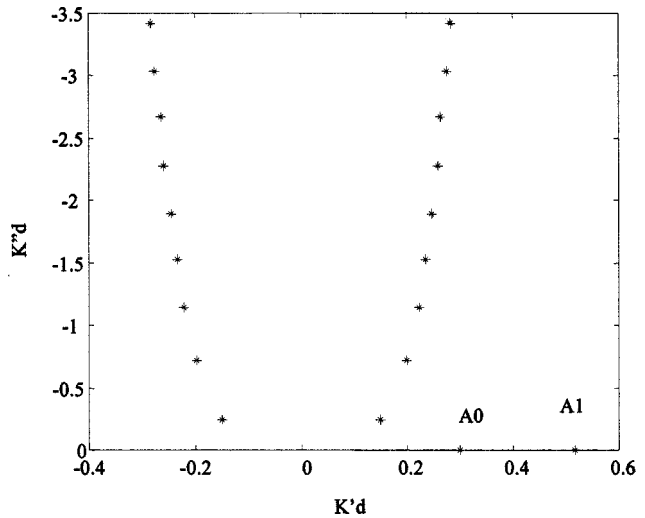


FIG. 3. The 20 first antisymmetric modes for $Fd = 4$ MHz mm in the complex wave number plane.

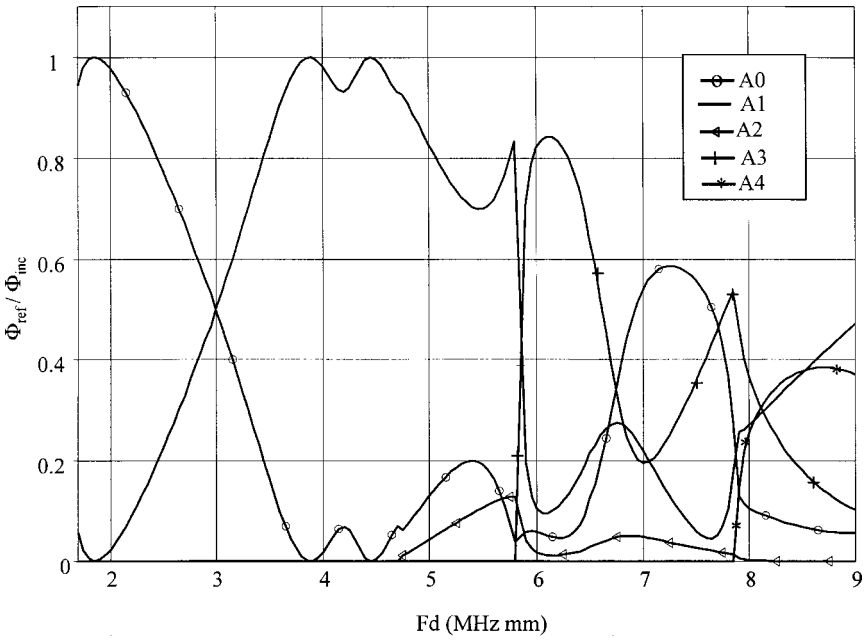


FIG. 4. Theoretical study. Repartition of the $A1$ incident wave energy among the reflected waves.

with a negative or a zero imaginary part K'' (damped in the negative x_1 direction) are taken into account. At this particular thickness-frequency product, there are only 2 real ($A0$, $A1$) and 18 complex roots (there is no imaginary solution).

The energy repartition of the incident $A1$ Lamb wave among the reflected modes is computed with a frequency-thickness step $\Delta Fd = 0.05$ MHz mm and plotted in Fig. 4 for $1.8 < Fd < 9$ MHz mm. There are three quite different zones. For the frequency-thickness products less than 2.3 MHz mm, the incident wave is almost completely converted into an $A0$ wave. Between 2.3 and 3.6 MHz mm, the trend is inverted and the $A0$ mode disappears to the benefit of the $A1$ mode. At each Fd value of a cutoff, an additional Lamb wave appears, so above $Fd = 5.85$ MHz mm the curves become intricate. In this condition, the balance between the incident and reflected energies is always achieved within 1% error. So, in the studied frequencies range, the $A1$ mode is almost completely converted in the first 20 real, imaginary and complex modes. We can neglect the contribution of the higher order modes and so limit the size of the matrix.

III. FINITE ELEMENT METHOD (FEM)

A. Excitation of a Lamb wave

In order to verify our theory, we use a finite element method. The FEM is used to solve static mechanical problems but can also be suitable for the study of wave propagation.[7] Computations are made with the finite element code ANSYS. The characteristics of the plate are identical to those used in the previous theoretical study. A two-dimensional study of the plate (25 mm long and 2 mm thick) with plane strain condition is performed (Fig. 5). This limited plate is regularly discretized in rectangular elements with 400 nodes in length and 20 in depth, so that there are at least 20 nodes per Lamb wave wavelength. The transient analysis is done with a 1.04×10^{-8} s integration time step (which satisfies the optimum time step equal to L/V if L is the element length and V is the phase velocity of the fastest Lamb wave in the studied frequency range.)

First, the wave conversion study implies that the incident wave should be unique at the selected frequency. For this purpose, the wave is numerically launched in the plate by imposing its theoretical longitudinal and shear displacements $u_1(x_3,t)$ and $u_3(x_3,t)$ during five cycles at a given frequency F to the 20 nodes of the plate extremity (Fig. 5).

In a first time we detail the results in the particular case of the $A1$ incident mode at a frequency-thickness product equal to 2.3 MHz mm. Of course this study is repeated at several other Fd and finally the energy reflection coefficients are obtained from $Fd = 2.3$ MHz mm to $Fd = 4.5$ MHz mm. In the case of the incident $A1$ mode at $Fd = 2.3$ MHz mm, the computed surface displacements are given in Fig. 6. From this B scan, we can get the phase v_p and the group v_g

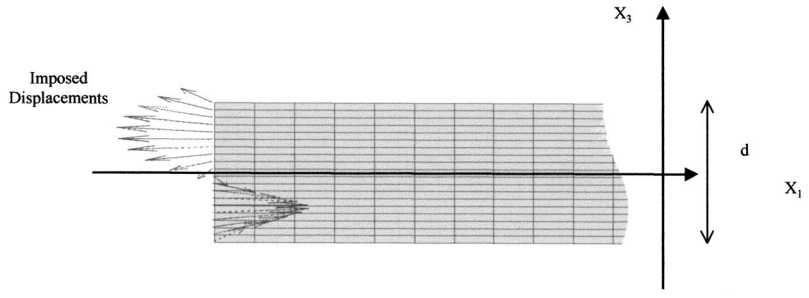


FIG. 5. Finite element method. Waves are numerically launched in the plate by imposing their theoretical longitudinal and shear displacements $u_1(x_3,t)$ and $u_3(x_3,t)$ to the 20 nodes of the plate extremity.

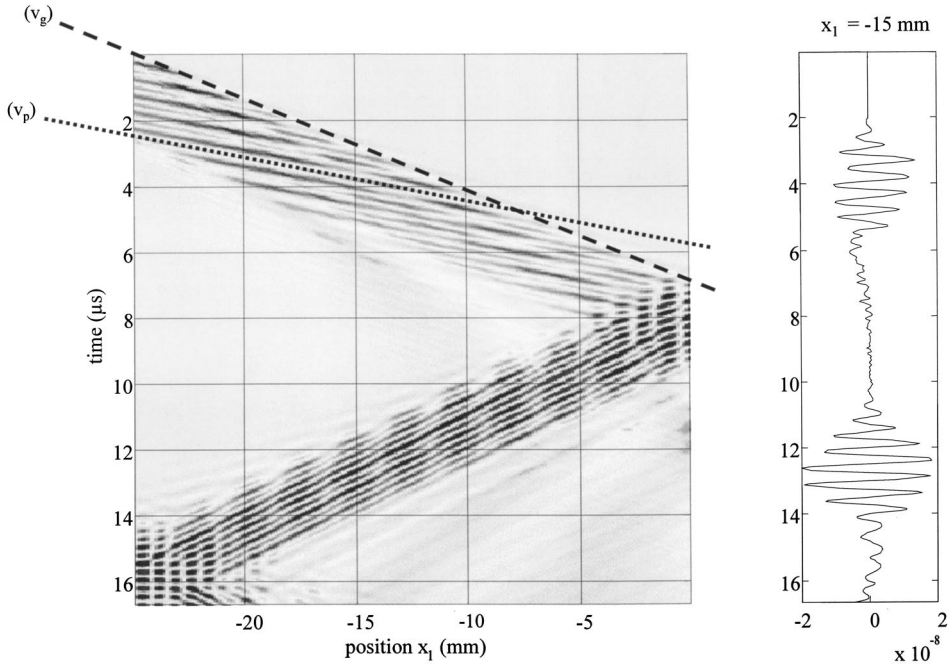


FIG. 6. Incident $A1$ wave and its reflection at the end of the plate. Spatio-temporal representation and time record at the location $x_1 = -15$ mm. The slopes $\Delta x_1/\Delta t$ of the lines give respectively the phase v_p and the group v_g velocities of the incident wave.

velocities of the incident wave. The slopes of the two dashed lines give respectively $v_p = 3650\ \mathrm{m\,s^{-1}}$ and $v_g = 7500\ \mathrm{m\,s^{-1}}$. These values are in a good agreement with the theoretical velocities of the $A1$ mode at $Fd = 2.3$ MHz mm. Moreover, considering the reflection of this wave, some amplitude oscillations mean that two or more reflected waves coexist. The propagation distance on the plate, however, is not long enough to separate the reflected modes in this spatio-temporal representation.

In order to identify the reflected waves, a spatial and a temporal fast Fourier transform are successively applied to the $u_3(x_1,t)$ signal,[8,9] providing the representation of the signal in the dual space (k,F). The modulus of the resulting signal $\psi(k,Fd)$ is plotted in gray level in Fig. 7. The positive (respectively negative) wave numbers are associated to waves propagating in the increasing x_1 direction (respectively opposite direction). Two Lamb waves are seen to be excited: the mode $A0$ at a Fd product below 0.5 MHz mm and the $A1$ mode over the frequency range 2–2.6 MHz mm. The $A0$ mode is probably generated from the transient phase of the burst. The $A0$ mode does not interfere with the $A1$ amplitude measurement. In fact the waves are excited in two quite different frequency ranges and at a given thickness-frequency product there is one incident Lamb mode at once. Therefore the technique used to generate numerically the waves is suitable for this study. In the case of the incident $A1$ wave, two waves are reflected: $A1$ mode and a strong con-

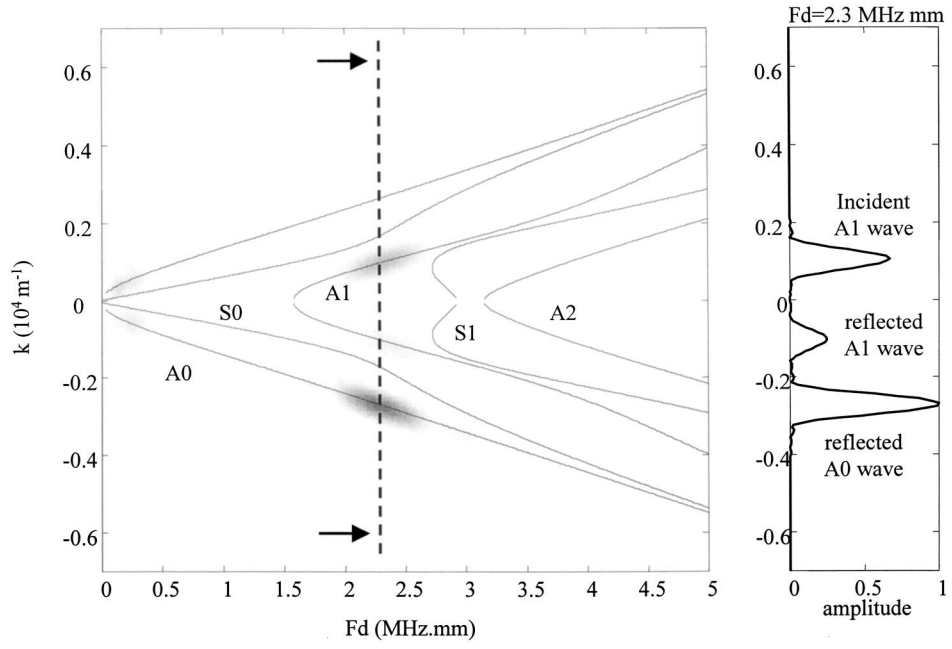


FIG. 7. Representation of the signal in the dual space (k,F). The solid lines are the theoretical dispersion curves.

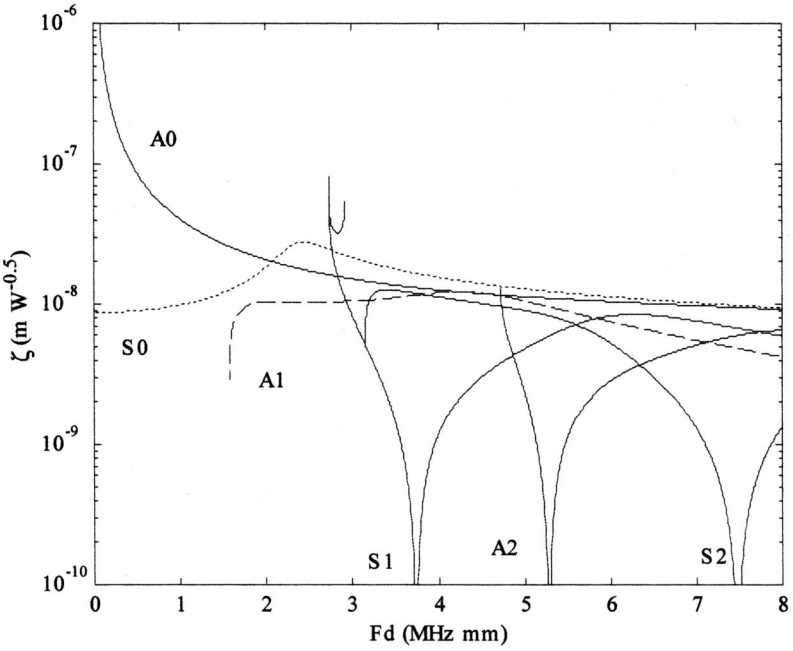


FIG. 8. Theoretical variation of coefficient ζ with frequency-thickness product for different Lamb modes. ζ relates the theoretical amplitude $|U_3|$ of the wave normal surface displacement to the square root of the power flow $|\phi|$.

verted $A0$ mode. Due to the symmetry of the scatterer, no symmetric mode is reflected.

In order to determine the power flow of the reflected waves, we must connect the measured normal displacement to the power.[10] This is done in the next section.

B. Results: Energy carried by the reflected Lamb waves at the end of a plate

We define an absolute coefficient $\zeta = |U_3(x_3 = \pm d/2)| / \sqrt{|\phi|}$ which relates the theoretical amplitude U_3 of the normal surface displacement of a Lamb wave to the square root of the power flow $|\phi|$. This theoretical coefficient is plotted in Fig. 8 for several Lamb modes and for thickness-frequency products ranging from 0 to 8 MHz mm.

From the representation of the signal in the dual space (k,F), we measure the displacement amplitudes $|\psi(k_n,F)|$ of the nth Lamb wave having a wave number k_n at a given frequency F $(n=1,\ldots,N)$. The energy of each Lamb wave is obtained from the measured amplitude of the displacement and the theoretical coefficient ζ:

$$\phi_n(F) = \left(\frac{|\psi(k_n,F)|}{\zeta_n(F)} \right)^2 .$$

Let us now consider the incident Lamb wave $A1$, at a given frequency F. It gives rise to N reflected Lamb waves. We obtain the energy reflection coefficient R_n as the ratio of the nth reflected wave energy ϕ_n to the $A1$ incident wave energy ϕ_{inc}:

$$R_n = \frac{\phi_n}{\phi_{\text{inc}}} .$$

The method (FEM computation and processing) is repeated at several other frequencies. The energy of the reflected modes is computed by 0.05 MHz mm step. Figure 9 represents the energy reflection coefficients over the frequency range 2.2–4.5 MHz mm. Theoretical energy reflection coefficients are plotted in solid line. The superimposed FEM results are in a good agreement with the theoretical ones.

IV. EXPERIMENTATION

A. Quasi-harmonic study

The experimental setup is described in Fig. 10. A generator delivers a 30-cycle toneburst at the frequency $F = 2.1$ MHz to a Panametrics emitting piezo-composite transducer. This transducer is wide frequency range with a 2.25 MHz central frequency. Lamb waves are generated by the wedge method. A Polytec laser vibrometer is used to measure the plate normal surface displacements along a line parallel to the incident Lamb wave propagation direction Ox_1. The displacement amplitudes are collected over 30 mm by 0.2 mm steps at the plate extremity. At each position, a 100-μs-long signal is registered (10 000 points). In order to im-

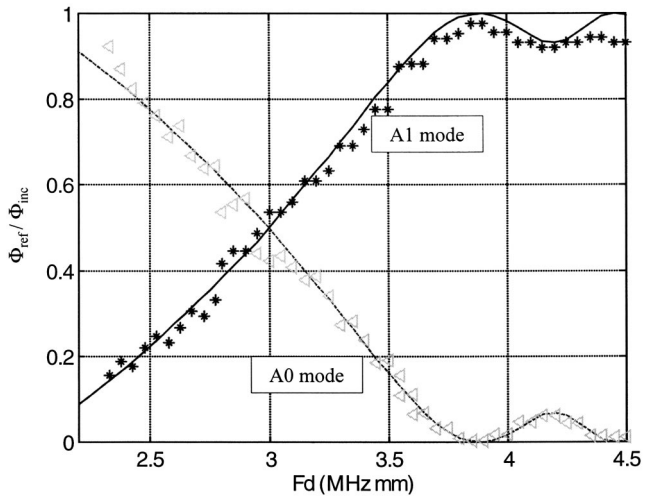


FIG. 9. Finite element method: energy reflection coefficients versus frequency-thickness in the case of an incident $A1$ wave.

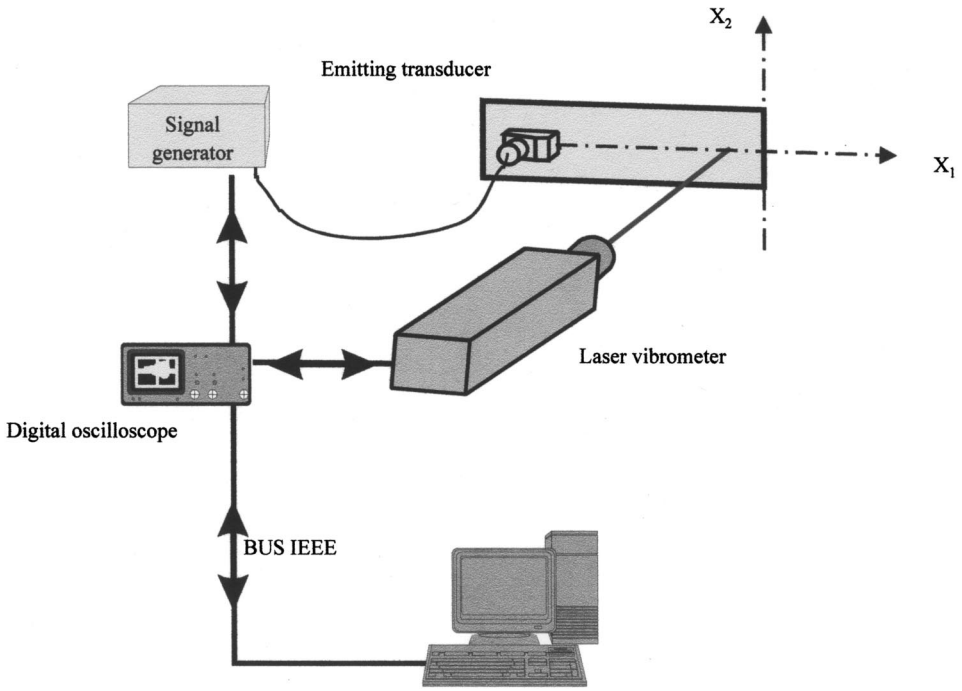


FIG. 10. Experimental setup.

prove the signal-to-noise ratio we perform an averaging over 1000 successive shots. The resulting B scan (Fig. 11), similar to those obtained with the finite element method, represents the normal displacement $u_3(x_1,t)$ versus time t and position x_1. Amplitude oscillations are observable on the reflected waves. This is due to the alternate constructive and destructive interference of two waves propagating with near velocities. On the spatial spectrum representation of the signal at $Fd=4.2$ MHz mm we verify that the incident wave is the $A1$ Lamb wave (Fig. 12). Moreover, this figure confirms the presence of two reflected waves: $A0$ and $A1$ modes. At this frequency, these two first antisymmetric modes only exist. By means of the ζ coefficient we obtain the energy carried by each waves: 47.9% of the incident energy in $A1$ wave and 2.2% in the $A0$ wave. These value should be compared to the theoretical ones: 4% of the incident energy converted in the $A0$ wave and 96% in the $A1$ wave. Obviously the determination of the reflected modes energy shows that the energy conservation is not verified: less than 50% of the incident wave energy is transmitted in the reflected modes. Several hypotheses can be given to explain the discrepancy between computation and experimentation. The main cause of the difference is that the $A1$ Lamb waves generated at the wedge/plate interface do not have a constant amplitude and it is obvious that the processing gives a mean value of the wave amplitude (reflected or incident) over a large distance of propagation. Indeed, the incident beam diverges during the propagation due to the angular aperture caused by the disc-shaped transducer.

The divergence of the wave during the propagation is

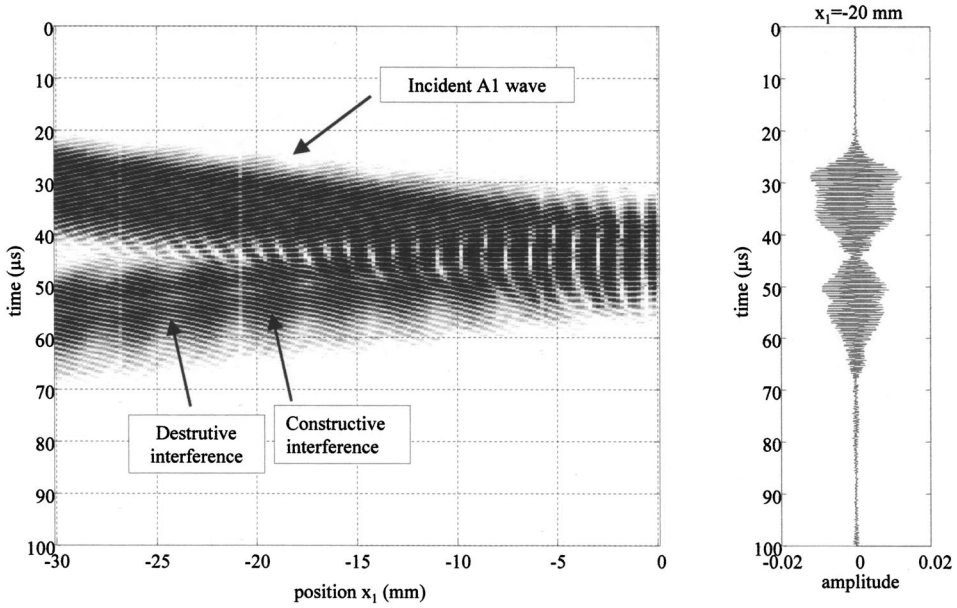


FIG. 11. Quasi-harmonic experimental study. Normal surface displacement $u_3(x_1,t)$ versus time and position x_1. Time recorded signal at the location $x_1=-20$ mm.

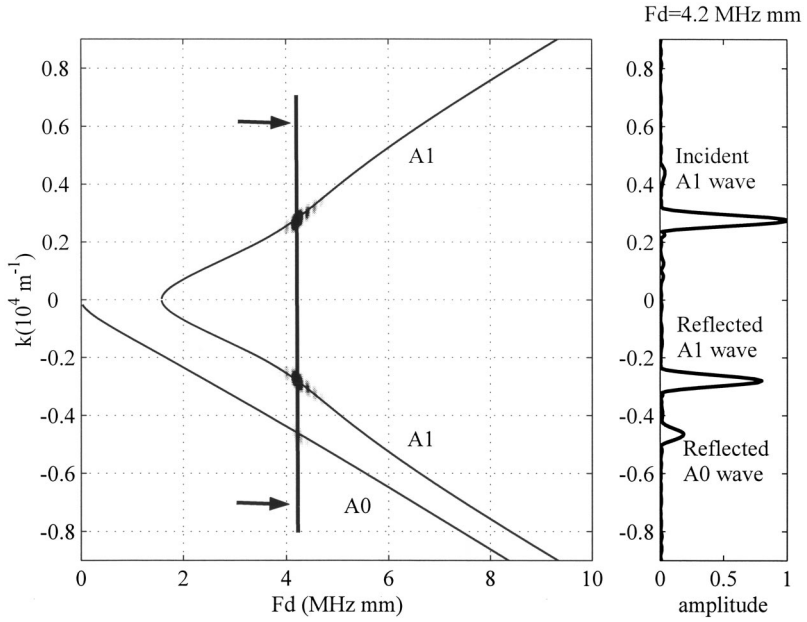


FIG. 12. Quasi-harmonic experimental study. Representation of the signal in the dual space (k,F). The solid lines are to the theoretical dispersion curves.

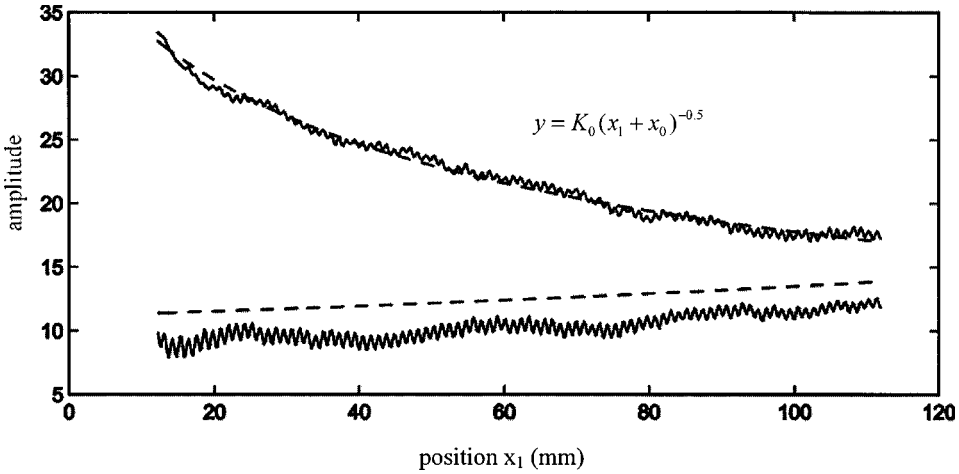


FIG. 13. Spatial behavior of the $A1$ wave amplitude at Fd=4.2 MHz mm. The superimposed dashed curve is the regression of the incident $A1$ wave amplitude.

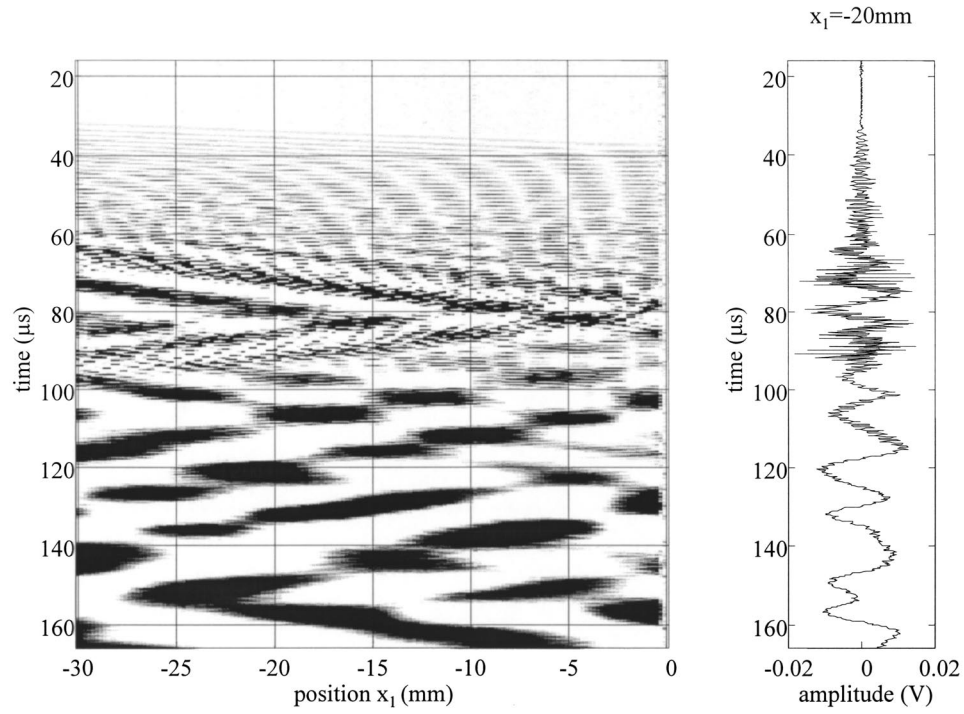


FIG. 14. Impulsional experimental study. Normal surface displacement $u_3(x_1,t)$ versus time and position x_1. Time recorded signal at the location $x_1 = -20$ mm.

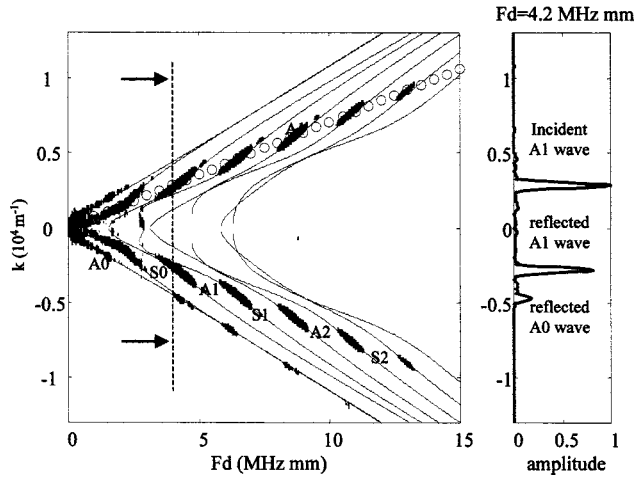


FIG. 15. Impulsional experimental study. Representation of signal in the dual space (k,F). The solid lines are the theoretical dispersion curves.

verified by the determination of the amplitude attenuation along x_1 direction. By means of the experimental setup described in Fig. 10, we measure the surface displacements along the transducer axis from $x_1=-120$ to $x_1=-10$ mm by 0.2-mm steps. On the resulting signal $u_3(x_1,t)$, we perform a temporal fast Fourier transform which gives $\chi(x_1,F)$ in the space-frequency representation. Let us consider the $A1$ mode at frequency $Fd=4.2$ MHz mm. In order to make an estimation of its amplitude attenuation during the propagation, a sliding window spatial Fourier transform is applied to the signal $\chi(x_1,F)$ at $F=2.1$ MHz. The 50-point rectangular window is translated along x_1 on the signal $\chi(x_1,F)$ and at each x_1 position a spatial FFT is performed on the windowed signal. In Fig. 13 we plot the amplitudes of the incident and reflected $A1$ waves. The superimposed dashed curve is the regression of the experimental peak amplitude of the incident $A1$ wave. The equation of this curve is $y=K(x_1+x_0)^{-1/2}$. This result agrees well with the theory of diffraction. Even if the uncertainty on the determination of reflected $A1$ wave amplitude is higher than the incident one, the wave continues to decrease as $1/\sqrt{x_1}$. In the next step it is assumed that the divergence is transmitted to the reflected wave so the converted wave decreases as the incident wave.

In addition to the beam divergence, some diffusion and attenuation of the Lamb waves occur in the material during the propagation (it is not negligible in this frequency range). The quasi-harmonic study allows us to obtain the energy repartition of an incident wave at a given frequency. In order to obtain the results on a large frequency range one must repeat the measurement at numerous frequencies and this way could be very expensive in time and in memory. A solution is to perform an impulsionnal study which covers a larger frequency bandwidth.

B. Impulsional results

The experimental setup is identical to the previous quasi-harmonic study and a pulse voltage (300 V) is applied on the emitter transducer. In Fig. 14 the resulting B-scan $s(x_1,t)$ describes the time evolution of the u_3 displacements collected by the laser vibrometer along a line parallel to the direction of propagation of the waves. In order to identify each Lamb wave, a spatial and a temporal fast Fourier transform are successively applied to the $s(x_1,t)$ signal, providing a new representation of the signal in the dual space (k,F). The 2d-FFT representation of the signal is shown in Fig. 15. Between $Fd=3.8$ and $Fd=4.8$ MHz mm, the incident wave is the $A1$ wave and there is no other incident wave with a significant amplitude. Therefore the study of the $A1$ mode conversion at the end of the plate is possible at these frequencies. At a given frequency, the reflected Lamb waves energies are deduced from the amplitudes by mean of the ζ coefficient previously defined. The results are corrected to take into account the beam divergence. Figure 16(a) shows the $A1$ incident energy repartition among the reflected modes over the frequency range 3.8–4.8 MHz mm. The $A1$ mode energy varies in an opposed trend to $A0$: the $A1$ energy decrease is compensated by the $A0$ energy increase. The variations of reflected wave energies are similar to the theoretical results [Fig. 16(b)] even if reflected energy is 80% of the incident one. In order to obtain a significant improvement in balance between incident and reflected energy, the experimentation must be done with a real plane wave and a 3D finite element computation must be performed.

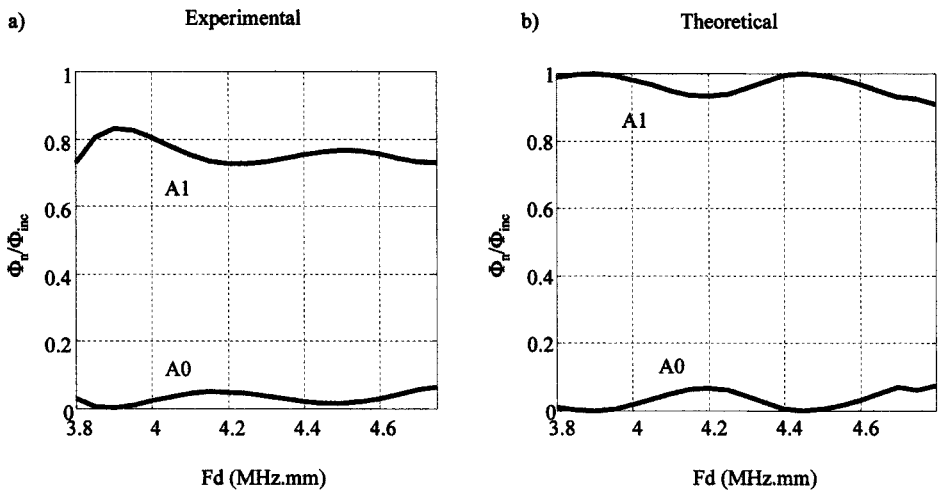


FIG. 16. Comparison between the experimental (a) and theoretical (b) energy reflection coefficient at the end of the plate (incident $A1$ wave).

V. CONCLUSION

Different studies of the reflection of the $A1$ Lamb wave at the free edge of a plate are performed in this paper. The FEM confirms our theoretical model with a good accuracy. Conversion modes are demonstrated and evaluated. An experimental verification of these conversions is achieved by vibrometer measurements of the wave displacement on the plate. The processing of the space data linked to the waves gives the reflection coefficient and the energy balance. We show that the incident wave energy is shared out among the reflected waves. These methods can be applied to study the Lamb wave conversion onto a wide variety of geometrical discontinuities. For example, in the case of a rib stiffened plate, we have made a precise nondestructive evaluation of the soldered joints by performing the measurement of the converted energy.[11]

[1] P. J. Torvik, ''Reflection of wave trains in semi infinite plates,'' J. Acoust. Soc. Am. **41**, 346–353 (1967).

[2] D. C. Gazis and R. D. Mindlin, ''Extensionnal vibrations and waves in a circular disk and a semi-infinite plate,'' J. Appl. Mech. **September**, 541–547 (1960).

[3] R. D. Gregory and I. Gladwell, ''The reflection of a symmetric Rayleigh-Lamb wave at the fixed or free edge of a plate,'' J. Elast. **13**, 185–206 (1983).

[4] Y. Cho and J. L. Rose, ''A boundary element solution for a mode conversion study on the edge reflection of Lamb waves,'' J. Acoust. Soc. Am. **99**, 2097–2109 (1996).

[5] S. Y. Zhang, J. Z. Shen, and C. F. Ying, ''The reflection of the Lamb wave by a free plate edge: visualization and theory,'' Mater. Eval. **46**, 638–641 (1988).

[6] M. Lowe and O. Diligent, ''Reflection of a fondamental Lamb modes from the ends of plates,'' *Review of Progress in Quantitative* NDE, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 2001), Vol. 20.A, pp. 89–96.

[7] D. N. Alleyne, M. J. S. Lowe, and P. Cawley, ''The reflection of guided waves from circumferential notches in pipes,'' J. Appl. Mech. **65**, 635–641 (1998).

[8] D. Alleyne and P. Cawley, ''A two-dimensional Fourier transform method for the measurement of propagating multimode signals,'' J. Acoust. Soc. Am. **89**(3), 1159–1168 (1991).

[9] L. Martinez, J. Duclos, and A. Tinel, ''Aspects of a damped surface wave in the Fourier diamond spaces. New surface wave analysis method (S.W.A.M.),'' J. Acoust. Soc. Am. **103**, 2901(A) (1998).

[10] B. Morvan, A. Tinel, and J. Duclos, ''Coupling of Lamb waves at a tee junction,'' IEEE International Ultrasonics Symposium, Lake Tahoe, NV, Vol. 1, pp. 565–568 (1999).

[11] B. Morvan, H. Duflo, J. Duclos, and A. Tinel, ''Lamb wave interaction with a welded rib,'' Acoust. Lett. **24**(6), 111–116 (2000).

Long-lasting stable cavitation

Jean-Louis Mestas,[a] Peter Lenz, and Dominique Cathignol
Institut National de la Santé et de la Recherche Médicale, Unité 556, 151 cours Albert Thomas, 69424 Lyon CEDEX 03, France

(Received 18 June 2001; accepted for publication 25 October 2002)

Stable cavitation is produced on the surface of a special target, i.e., an abrasive foil with a grain size of about 15 μm, insonified by a 473 kHz focused beam. Cavitation bubbles are first created by a pressure of about 2 MPa. Progressive reduction of the pressure to about 100 kPa leads to a state of stable cavitation characterized by strong stable emission of the half-order subharmonic. This state can be maintained for five hours if the defining parameters (position and pressure) are optimized and constant to within a few percent. There is strong evidence for the presence of ''latent'' cavitation bubbles, which can persist for a few minutes without being excited. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1538198]

PACS numbers: 43.35.Ei, 43.25.Yw [ADP]

I. INTRODUCTION

A difficulty encountered in experiments on stable cavitation is the short time during which cavitation bubbles can be studied. Several proposals have been made in the past to increase this time. Mostly bubbles produced elsewhere are trapped at the surface of different objects. These may be the tip of a wire,[1] a thin wire coated or not with a film of paraffin wax,[2,3] a hydrophone,[4,5] a tissue formed by glass–wool fibers,[4] or a porous surface.[6] The latter three configurations allow for the study of multiple bubbles, which is advantageous because the signal intensity is enhanced. Another modality consists of concentrating bubbles in the antinodes of standing waves.[7–9]

None of these methods makes observation times longer than a few minutes possible. In the present study a method is described that produces multiple cavitation bubbles and thereafter traps them virtually indefinitely. It provides ease of observation and the possibility of characterizing the domain of stability of the bubbles.

II. MATERIALS AND METHODS

The experimental setup is shown in Fig. 1. The device, composed of an ultrasound emitter, a target, and an ultrasound receiver, is placed in a container filled with tap water, the oxygen content of which is 7–8.5 mg/l (Oxygen meter WTW Oxi 340). A silicone absorber (Silastic 3120 Dow Corning) facing the emitter reduces reflections.

The ultrasound emitter is a PZT transducer (Channel Industries Inc. C5500) of 100 mm diameter and 100 mm curvature radius. It is equipped with a transformer so as to exhibit a virtually real impedance [(47+j 0.06) Ω] at the resonance frequency $\nu_0 = 473$ kHz. Its focal volume, corresponding to pressure values above 50% of the pressure at the focal point, has a depth of 35 mm and a diameter of 8 mm. A sinusoid generator tuned to 473 kHz (Hewlett Packard 33120 A), an attenuator and an amplifier (Kalmus LA 200H CE) can provide a continuous power between 7 mW and 160 W; this corresponds to a pressure between 35 kPa and 5.3 MPa at the focal point. For great pressure values nonlinear propagation appeared but did not influence the measured parameters. The voltage delivered by the generator is called the excitation E and is given in dBm.

The peak-to-peak amplitude of the amplifier output voltage is called U, measured in volts. The power absorbed by the emitting transducer, measured with a watt meter (Rhode & Schwartz NAP 392.4017.04) is called P (W).

With a target at or close to the focal point, the reflected wave influences the power supply. In order to determine the pressure at the target without a hydrophone, a specific study has been carried out.[10] It consisted of measuring, as a function of U and P, the maximal pressure present at targets that have different reflectivities but are located at one and the same point within the focal zone. It turns out that the pressure at the targets is well approximated by the equation

$$p(\mathrm{kPa}) \approx 180\sqrt{P(\mathrm{W})} + 12U(\mathrm{V}). \quad (1)$$

The receiving transducer, which is identical to the emitter, has its axis oriented perpendicularly to that of the latter. The received signal is fed through a wide band isolation transformer, amplified (NF Electronic Instruments BX31) and then sent either to a spectrum analyzer (Hewlett Packard ESA L 1500 A) or to an oscilloscope (Lecroy 9310 AL). The spectrum analyzer is used in two different modes. In the first mode the examined frequency range is 0.1–1.1 MHz, the bandwidth 10 kHz, and the duration of analysis 50 ms. The spectrum of the received signal is displayed, in particular, the lines $\nu_0/2$, ν_0, $3\nu_0/2$, and $2\nu_0$. To reduce noise, ten spectra are averaged. In the second mode only the line $\nu_0/2$ is examined within a 10 kHz band by measuring 400 successive intensity values during 50 ms. The mean and the standard deviation of these values are called $\mu_{\nu_0/2}$ and $\sigma_{\nu_0/2}$, respectively.

The target is an abrasive foil (thickness 0.1 mm, diameter 10 mm) whose usual destination is polishing optical fibers (type 15). Figure 2 shows the structure of its surface, suggesting that the type number stands for the grain size in micrometers. The target is glued to the center of a 0.06 mm

[a]Author to whom correspondence should be addressed. Electronic mail: mestas@lyon.inserm.fr

 0001-4966/2003/113(3)/1426/5/$19.00

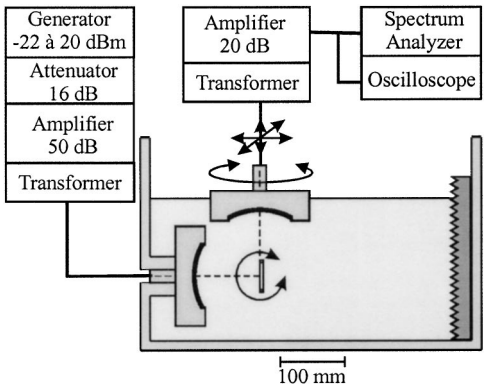


FIG. 1. Experimental setup.

thick plastic foil, which is stretched over a 50 mm diam PVC ring. All these elements are fixed to the receiving transducer so as to place the target at its focal point.

The receiving transducer is mounted on three translation stages. So the target can be positioned in the focal zone of the emitter; it is also oriented perpendicular to the emitter axis (two angular adjustments of the PVC ring).

III. RESULTS

A. Initiation of cavitation

With weak excitation ($p \approx 60$ kPa) and a target carrying relatively few bubbles, no cavitation is observed and the spectrum exhibits only a line corresponding to the emission frequency [Fig. 3(a)]. To initiate cavitation, a pressure between 1.5 and 3 MPa is applied to the target during about 10 s. Visible bubbles are produced in the conical field of the emitting transducer. They gather on equidistant, concentric shells and can jump from one shell to another in the direction of the focal point or escape toward the surface. The spectrum exhibits abundant noise and two principal lines at ν_0 and $2\nu_0$ [Fig. 3(b)]. Lowering p leads to the disappearance of visible bubbles. The noise level is reduced and two more lines appear at $\nu_0/2$ and $3\nu_0/2$ [Fig. 3(c)]. Figure 3(d), obtained with $p \approx 100$ kPa, shows an increase of the lines at $\nu_0/2$ and $3\nu_0/2$ and a decrease of the lines at ν_0 and $2\nu_0$. Reducing p progressively and adjusting the target's position on the emitter axis (see below) make it possible to attain an ''operating point,'' which is defined by two criteria: (1) the intensity of the line $\nu_0/2$ is high and stable even in a single (not averaged) spectrum [Fig. 3(e)]; (2) the received signal (Fig. 4), composed of frequencies ν_0 and $\nu_0/2$, shows no fluctuation.

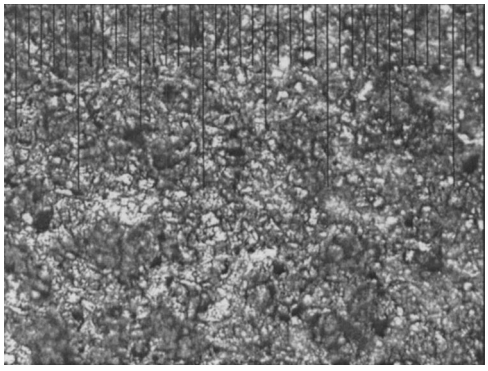


FIG. 2. Microscopic view of the target (10 μm/div).

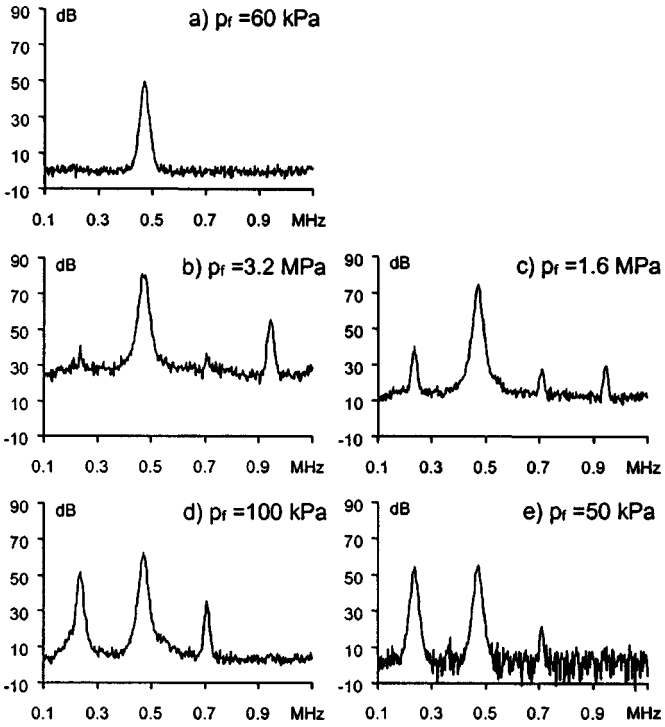


FIG. 3. Spectra of the received signal before and after the creation of oscillating bubbles, depending on the applied pressure (ordinate: intensity; abscissa: frequency).

B. Limits of operation

A pressure decrease of only a few kilopascals and/or an axial shift of the target by as little as 0.2 mm make the line $\nu_0/2$ disappear. Therefore the parameters that determine the operating point have to be chosen very carefully. Observing the dependence of $\mu_{\nu_0/2}$ and $\sigma_{\nu_0/2}$ on the target's position and on E allows one to define the ''operating domain.''

The influence of E is shown in Fig. 5. The operating domain corresponds to particularly low values of $\sigma_{\nu_0/2}$. The abrupt decrease of $\mu_{\nu_0/2}$ marks the lower limit even more precisely. For constant E Figs. 6 and 7 show $\mu_{\nu_0/2}$ and $\sigma_{\nu_0/2}$ as a function of the target's position on the emitter axis, near the focal point. In Figs. 6(a), (b), several distinct operating domains are seen, separated from one another by 1.6 mm intervals that correspond to half the emission wavelength. [Figure 6(c) will be explained in Sec. IV.] Figure 7 is a more detailed view of the principal operating domain around the focal point. It is clearly delimited by the abrupt decrease of $\mu_{\nu_0/2}$ and the increase of $\sigma_{\nu_0/2}$. Figure 8 shows the dependence of $\mu_{\nu_0/2}$ and $\sigma_{\nu_0/2}$ on the target's position on a line

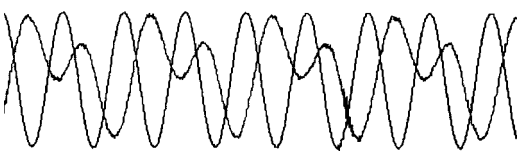


FIG. 4. Supply voltage ($\nu_0 = 473$ kHz) and received signal (containing harmonics) as a function of time. Arbitrary ordinate scales.

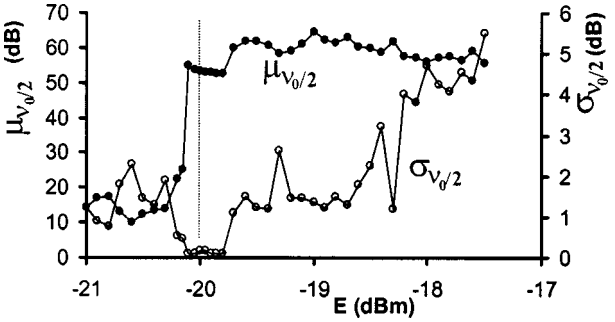


FIG. 5. Dependence of the received signal on excitation.

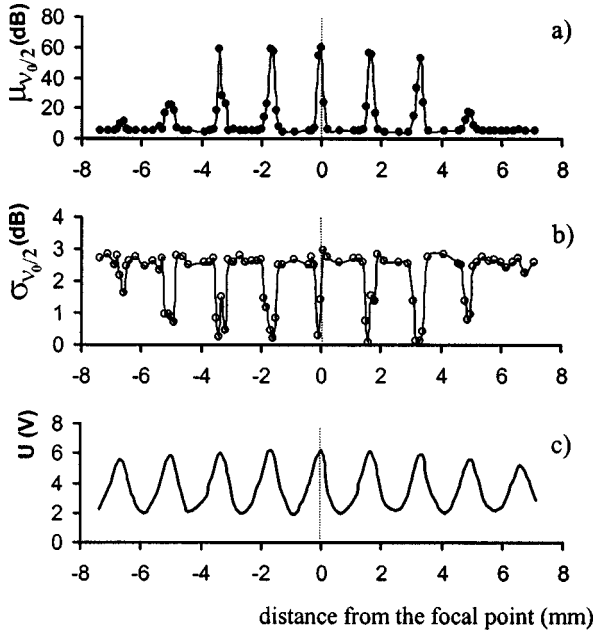


FIG. 6. Dependence of the received signal (a), (b) and of the supply voltage (c) on the target's axial position.

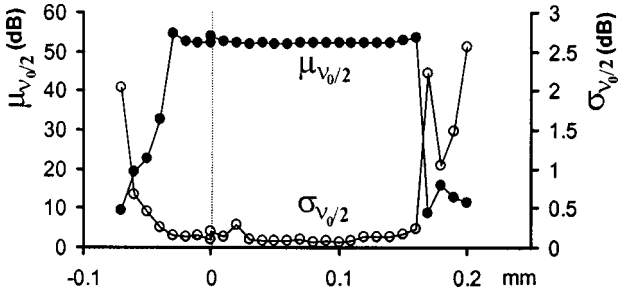


FIG. 7. Dependence of the received signal on the target's axial position (central operating domain).

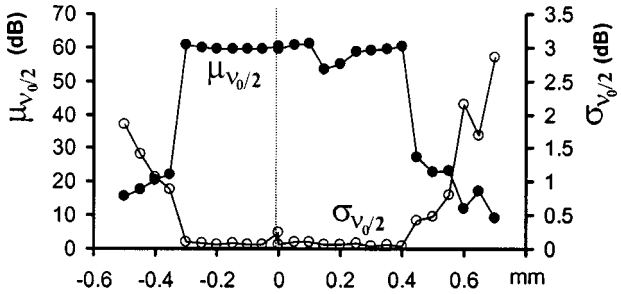


FIG. 8. Dependence of the received signal on the target's position on a line perpendicular to the axes of both transducers.

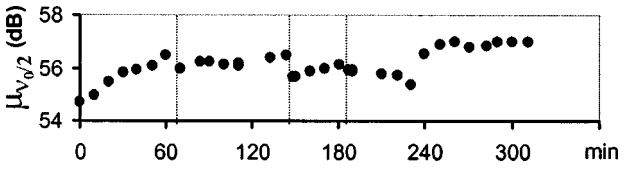


FIG. 9. Long-term stability of the received signal.

perpendicular to the axes of both transducers. Again, the operating domain is clearly delimited but larger than the one of Fig. 7.

In all these experiments measurements are carried out, starting from an operating point (which is indicated by a vertical line in the figures) and then moving to higher and lower abscissa values. It is worth mentioning that the $\mu_{\nu_0/2}$ and $\sigma_{\nu_0/2}$ values corresponding to the initial operating point change very little during this procedure, although the system has momentarily left the operating domain.

Interestingly, $\mu_{\nu_0/2}$ increases slightly (1–2 dB) close to the margins of the operating domain (Figs. 5–7). The increase is quite small but is always observed in experiments of these type.

C. Long-term stability

To observe the long-term stability of the system, it is placed in the principal operating domain and $\mu_{\nu_0/2}$ and $\sigma_{\nu_0/2}$ are measured during several hours (Fig. 9). $\mu_{\nu_0/2}$ increases slowly by 1–2 dB until $\sigma_{\nu_0/2}$ increases too, but abruptly. At this moment E has to be increased by 0.05 dBm in order to restore the initial situation. In Fig. 9 a vertical line indicates every increase of E. E initially equals -20.1 dBm and attains -19.95 dBm after 5 h of uninterrupted operation. $\sigma_{\nu_0/2}$ lies between 0.06 and 0.09 dB.

D. Reproducibility

The system is repeatedly placed in an operating domain. Between successive experiments, there is a waiting time of at least 5 min, requiring reinitiation. The position of the target is unchanged. Figure 10 shows that E varies by 0.3 dBm only. The variations of $\mu_{\nu_0/2}$ are stronger and $\sigma_{\nu_0/2}$ lies between 0.06 and 0.13 dB.

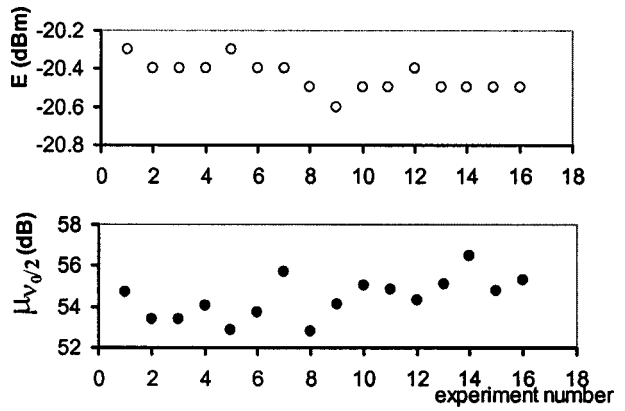


FIG. 10. Reproducibility of the received signal.

IV. DISCUSSION

In a liquid insonified with a frequency ν_0 cavitation manifests itself by the appearance of harmonics, in particular, the one at $\nu_0/2$.[11] It is attributed to bubble oscillation. Stable cavitation is obtained when the bubbles attain a radius corresponding to resonance at $\nu_0/2$.[12,13] In this study, bubbles are produced from seeds or from microbubbles by the action of high pressure (no effect is found in water containing only 3 mg/l of oxygen, even with a pressure of 3 MPa). To observe the phenomena described here, the production of bubbles alone is insufficient, however. The essential point is their fixation on a target having particular characteristics.

A. Characteristics of the target

The surface roughness of the target plays a decisive role for the production and trapping of resonant bubbles. This is shown by comparing a smooth plastic foil and abrasive foils of different grain sizes (types 3, 15, and 32). With a smooth foil, a pressure of more than 1 MPa is necessary to produce a signal at $\nu_0/2$. This implies transient (inertial) cavitation, which is not the object of this study. For the three rough targets $\mu_{\nu_0/2}$ is similar if E is similar. However, the signal is unstable ($\sigma_{\nu_0/2}>1.5$ dB) with a type 3 abrasive foil and it persists little with a type 32 foil. Only the type 15 foil provides simultaneously short-term and long-term stability of the received signal. One may try to explain these findings in the following manner.

The abrasive particles are randomly distributed on the target surface and their dimensions (in μm) correspond to the type numbers of foils. The spaces between particles, where bubbles are trapped, have about the same size. In the present case the diameter of bubbles that exhibit resonance at $\nu_0/2$ is approximately 28 μm (calculated according to Ref. 5). Thus, on a type 3 foil, the free space is insufficient for bubbles to oscillate in a unique mode. The frequent change of modes causes high values of $\sigma_{\nu_0/2}$. A type 32 foil exhibits wider spaces. Bubbles located there remain relatively mobile and therefore escape easily from the insonified zone. As a result, a stable but little persistent signal is observed. The spaces of a type 15 foil are such that bubbles are trapped efficiently while oscillating freely.

B. Influence of target reflectivity

In Fig. 6, five major operating domains appear. They are attributed to pressure maxima. This hypothesis is confirmed by measuring the pressure with a hydrophone having very low reflectivity (25 μm membrane GEC Marconi Y-33-7611) that replaces the target (Fig. 11, thin line). As a reference, the pressure has also been measured in the pulsed mode (repetition rate: 1 kHz; pulse length: 42 μs) with the same hydrophone. The distance between consecutive maxima equals half the emission wavelength (1.6 mm). To explain the presence of these maxima and minima one can suppose that the pressure wave, partly reflected by the hydrophone or the target, produces an electric signal in the emitter and thus influences the functioning of the emitter electronics, depending on the axial position of the reflector.[14] The pressure and the voltage U are modulated in a similar way [Fig. 6(c)]. This

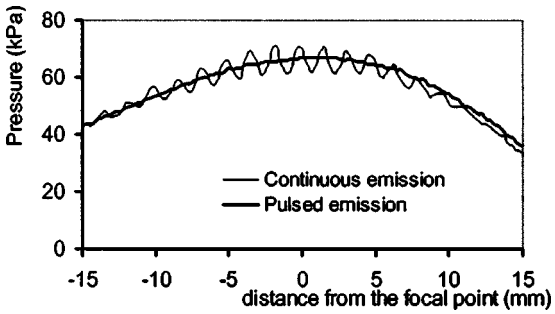


FIG. 11. Pressure amplitude measured on the emitter axis (emitter on the right).

effect is of practical interest because it allows one to place the target close to an operating domain simply by searching the maximum of U. To quantify this amplitude modulation the ratio of neighboring extreme values of U is determined near the focal point. With the hydrophone as a reflector, the ratio equals 1.2. The type 15 target has higher reflectivity and consequently produces a higher ratio (1.4). The ratio even attains the value 3.2 if stable cavitation occurs because the presence of numerous bubbles increases reflectivity (Fig. 6).

The electric reaction, produced by the reflection of the pressure wave off the target, also causes an alteration of the absorbed power for a given excitation. For $E=-19$ dBm, P equals 19 mW ($U=3.0$ V) in the absence of any reflector, but 33 mW ($U=5.8$ V) with a type 15 target functioning in its principal operating domain. The corresponding values of p [Eq. (1)] are 58 and 102 kPa, respectively.

This last value is not far from the threshold pressure of stable cavitation, i.e., 89 kPa in water, which in turn is proportional to the total damping coefficient of radial bubble motion.[12] The estimation of this coefficient could be an interesting application of the technique described here.

C. Limits of operation

As mentioned above, operating domains exist around pressure maxima (Fig. 6). Due to the approximately sinusoidal shape of the curve representing pressure (Fig. 11), the relative decrease of pressure at the edges of the operating domain (Fig. 7) can be calculated. One obtains 2%. Similarly, in Fig. 5, the width of the operating domain corresponds to a variation of E and therefore to a variation of pressure. Equation (1) gives a value of 3%.

These estimations show the narrowness of the pressure range within which the phenomenon can be observed. It is therefore not surprising that the slightest mechanical or thermal perturbation can throw the system out of its operating domain. When this occurs, a reinitiation is not necessarily required. Indeed, if the interruption does not last for more than one or two minutes, the system can be brought back to its operating domain by adjusting E or the target's position on the emitter axis. In particular, during the scanning necessary to obtain the results shown in Figs. 5–8 the system leaves the operating domain and reenters it easily. These observations suggest the present of ''latent'' bubbles that means

ones that have been created on the target but do not oscillate outside an operating domain.

Figures 5 and 7 show an interesting detail: At the edges of the operating domain $\mu_{\nu_0/2}$ increases slightly (1–2 dB) just before its abrupt decrease if the operating domain is limited by a decrease of pressure (the left side in Fig. 5 and both sides in Fig. 7). One is tempted to conclude that the increase of $\mu_{\nu_0/2}$ and the decrease of pressure are linked together; the mechanism, however, remains unclear.

During long-term operation (about 1 h) $\mu_{\nu_0/2}$ tends to increase by 1–2 dB (Fig. 9) until $\sigma_{\nu_0/2}$ increases abruptly. The system can be brought back into its initial state by a slight increase of E, which indicates a previous pressure decrease. Not only can this be considered as a third operating limit, but it also implies that the mechanism mentioned above is again at work. In the case of long-term observation, the pressure decrease may be due to a loss of bubbles (and consequently a decrease of the target's reflectivity) or to thermal drift of the electronics. It should be stressed that the duration of 5 h observed in Fig. 9 was imposed by the operator, not by the system. The experiment could have been carried on much longer.

D. Reproducibility

Here the variability of $\mu_{\nu_0/2}$ has no geometrical cause since the position of the target is unchanged. It is not the result of the varying duration of the strong initiating pressure either, if this duration is sufficient (about 10 s). An explanation can rather be found in the variability of the search for the operating point, because the progressive reduction of E is hardly reproducible. Another cause may be the fact that the operating point is generally situated near the edge of the operating domain, which means in a region that is characterized by a variation of $\mu_{\nu_0/2}$ of 1–2 dB (Figs. 5 and 7).

ACKNOWLEDGMENT

We would like to thank Mr. A. Mathias for his technical assistance.

[1] A. D. Phelps and T. G. Leighton, ''The subharmonic oscillations and combination frequency subharmonic emissions from a resonant bubble: their properties and generation mechanisms,'' Acustica **83**, 59–66 (1997).
[2] A. D. Phelps and T. G. Leighton, ''High resolution bubble sizing through detection of the subharmonic response with two frequency excitation technique,'' J. Acoust. Soc. Am. **99**, 1985–1992 (1996).
[3] T. G. Leighton, R. J. Lingard, A. J. Walton, and J. E. Field, ''Acoustic bubble sizing by combination of subharmonic emissions with imaging frequency,'' Ultrasonics **29**, 319–323 (1991).
[4] E. A. Neppiras, ''Subharmonic and other low-frequency emission from bubbles in sound-irradiated liquids,'' J. Acoust. Soc. Am. **46**, 587–601 (1969).
[5] E. A. Neppiras, ''Subharmonic and other low-frequency signals from sound-irradiated liquids,'' J. Sound Vib. **10**, 176–186 (1969).
[6] W. L. Ryan and E. A. Hemmingsen, ''Bubble formation at porous hydrophobic surfaces,'' J. Colloid Interface Sci. **197**, 101–107 (1998).
[7] L. A. Crum and A. Prosperetti, ''Nonlinear oscillations of gas bubbles in liquids: An interpretation of some experimental results,'' J. Acoust. Soc. Am. **73**, 121–127 (1983).
[8] L. A. Crum, ''Measurements of the growth of air bubbles by rectified diffusion,'' J. Acoust. Soc. Am. **68**, 203–211 (1980).
[9] A. Eller, ''Force on a bubble in a standing wave,'' J. Acoust. Soc. Am. **43**, 170–171 (1968).
[10] J. L. Mestas, P. Lenz, and D. Cathignol, to be published elsewhere.
[11] P. W. Vaughan, ''Investigation of acoustic cavitation thresholds by observation of the first subharmonic,'' J. Sound Vib. **7**, 236–246 (1968).
[12] A. Eller and H. G. Flynn, ''Generation of subharmonics of order one-half by bubbles in a sound field,'' J. Acoust. Soc. Am. **46**, 722–727 (1969).
[13] F. R. Young, *Cavitation* (McGraw-Hill, London, 1989), pp. 117–119.
[14] D. R. Bacon, ''Measurements on a specific acoustic pulse Part 2: Measurements,'' in *Output Measurements for Medical Ultrasound*, edited by R. C. Preston (Springer-Verlag, Berlin, 1991), Chap. 7, pp. 122–123.

Characterization of an optical multilayer hydrophone with constant frequency response in the range from 1 to 75 MHz

Volker Wilkens[a)]
Physikalisch-Technische Bundesanstalt, Fachlabor Ultraschall, Bundesallee 100, 38116 Braunschweig, Germany

(Received 29 July 2002; revised 13 December 2002; accepted 17 December 2002)

The performance of an optical multilayer hydrophone for ultrasound measurements is investigated both in theoretical terms and experimentally. The optical measurement system comprises a thin high-finesse dielectric interference filter structure that is deposited onto a plane glass plate. An incident acoustic pressure wave deforms the layer system, and the induced variation of the optical reflectance is determined. Applying the concept of an optical off-axis detection scheme offers good sensitivity and a simple and low-cost setup. A primary interferometric calibration technique is applied to experimentally determine the pressure–voltage transfer function in the range from 1 to 75 MHz. Within the measurement uncertainty a constant transfer factor is obtained for the whole frequency range. Measurements of broadband ultrasound pulses are influenced neither by acoustic resonances of the very thin sensing element nor by diffraction phenomena that are known to cause waveform distortions in small probe hydrophone measurements. High temporal and spatial resolution is combined with high durability of the probe, which is why the optical multilayer hydrophone is well suited for use as a reference for secondary hydrophone calibration. [DOI: 10.1121/1.1553457]

PACS numbers: 43.38.Zp, 43.35.Yb, 43.30.Yj [YHB]

I. INTRODUCTION

Optical measurement systems have proved to be useful tools for the detection of ultrasound, in particular, in the high-frequency range. The use of thin optical dielectric layers as a sensor element in both a fiber-optic hydrophone and an optical detection array configuration has been reported on previously.[1,2] Acoustic measurement here is based on the elastic deformation of the multilayer system by an incident ultrasound wave. The thin layer system is designed to yield a large change in optical reflectance at the operational light wavelength produced by this mechanical deformation. The change in reflectance is detected using a simple optical setup, and high temporal and spatial resolution of the ultrasound measurement can be achieved. The frequency response of the fiber-optic version of the multilayer hydrophone is strongly influenced by the acoustic properties of the fiber end onto which the layer system is deposited. It can be understood as the result of the superposition of longitudinal, edge diffraction, and lateral waves with a resonant vibration mode of the fiber body representing an elastic rod.[3] The pressure waveforms of broadband ultrasound pulses measured with such a fiber-optic multilayer hydrophone are strongly distorted as a result of the nonuniform frequency response. If a probe is used that comprises a similar layer system deposited onto a plane glass plate with a diameter larger than the lateral sound field dimension, such distortions do not occur since edge and resonance effects are avoided.[2] In addition to the uniform frequency response, this sensor concept basically offers high temporal and spatial resolution and high durability of the probe, which is why the system is well suited for use as a reference detector for secondary calibrations of hydrophones and fiber-optic ultrasound sensors, especially for the high-frequency range.

In this paper, the concept of an optical off-axis detection scheme for the multilayer hydrophone is investigated theoretically and by experiments. It offers both, good sensitivity and a simple and low-cost setup, and adjustment to the operating point of the optical multilayer resonator can be achieved by variation of the illumination angle.[4,5] A similar adjustment technique has been demonstrated to be useful for high-finesse etalon-type ultrasound sensors[6] and more recently for low-finesse polymer film Fabry–Perot-type ultrasound sensors.[7] The acoustic performance of the multilayer hydrophone is experimentally analyzed. The characterization given comprises an estimation of the lateral resolution, the frequency response determined by a primary interferometric calibration in the range from 1 to 75 MHz, the sensitivity, i.e. the minimum detectable sound pressure level of the present system, the directional response, and the stability.

II. SENSOR DESIGN, PRINCIPLE OF OPERATION, AND THEORETICAL DESCRIPTION

The sensor element of the optical multilayer hydrophone consists of a dielectric layer system acting as a microinterferometer deposited on a plane glass plate (Fig. 1, insert). The high-finesse dielectric interference filter structure is of a design well suited for use as a pressure sensing element at ultrasound frequencies.[1,5] Two high-reflection subsystems, both consisting of several $\lambda_D/4$ layers, are connected by a central $\lambda_D/2$ spacer layer, where λ_D denotes the optical design wavelength at which the interference filter offers maximum transmission. This structure combines a high optical

a)Electronic mail: Volker.Wilkens@ptb.de

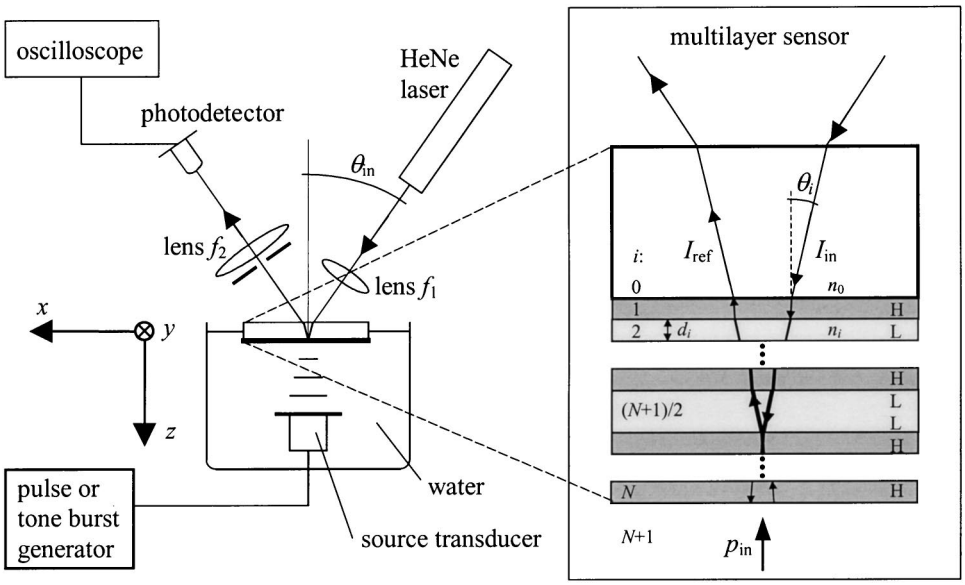


FIG. 1. Optical multilayer hydrophone setup; insert: multilayer system deposited onto a plane glass substrate; $i=0$: substrate, $i=1,...,N$: dielectric layers, $i=N+1$: surrounding water, N: total layer number, H, L: high-index and low-index optical $\lambda_D/4$ layer, respectively, p_{in}: incident pressure wave, I_{in}: incident, and I_{ref}: reflected optical intensity, n_i, θ_i: refractive index and light propagation angle, respectively, in the ith medium.

finesse of the Fabry–Perot interferometer with a very small overall thickness, and good acoustic sensitivity together with a large detection bandwidth is obtained.

The principle of ultrasound measurement is based on the elastic deformation of the layer system by an incident pressure wave p_{in} and the measurement of the induced change in optical reflectance ΔR by an optical detection scheme (Fig. 1). The optical arrangement for a single element detector can be considerably simplified if the oblique incidence of light is utilized instead of perpendicular illumination in the setup previously described.[2] Only very few optical components comprising the laser source, two lenses, the sensor probe, and a photodetector are necessary. Focusing of the laser beam yields high spatial resolution of the ultrasound measurement. Since backreflection into the laser source does not occur due to the off-axis detection scheme, optical isolation is not necessary. Backreflections from the rear of the substrate are laterally separated from the light reflected at the layer system and can easily be masked to avoid parasitic interference. Furthermore, a beamsplitter needs not to be inserted for the detection of the light reflected at the layer system, which offers highly efficient use of light, and a laser source with comparatively low output power is sufficient to achieve good sensitivity of the optical hydrophone. For the monochromatic light source here a HeNe laser is preferred to a laser diode because mode competition noise typical of laser diodes would give reason to increased intensity noise in the reflected part of light.[5]

For a theoretical description the system of N dielectric layers is assumed to be optically nonabsorbing, homogeneous, and isotropic. The optical reflectance R of the system for oblique incidence of light with wavelength λ can then be calculated using the matrix formalism for optical multilayer systems,[8] where the single ith layer with the physical thickness d_i, refractive index n_i, and angle of light propagation θ_i is described by its angle-dependent phase thickness:

$$\phi_i = \frac{2\pi}{\lambda} n_i d_i \cos\theta_i, \quad i=1,...,N, \tag{1}$$

and its characteristic matrix:

$$\mathcal{M}_i = \begin{bmatrix} \cos\phi_i & \dfrac{\sqrt{-1}\sin\phi_i}{\gamma_i} \\ \sqrt{-1}\,\gamma_i \sin\phi_i & \cos\phi_i \end{bmatrix}, \quad i=1,...,N. \tag{2}$$

Here, γ_i denotes the optical admittance of the layer that depends on the state of polarization of the incident light:

$$\gamma_i = \sqrt{\frac{\epsilon_0}{\mu_0}}\, n_i \cos\theta_i, \quad \text{for perpendicular } (s) \text{ polarization;}$$
$$\gamma_i = \sqrt{\frac{\epsilon_0}{\mu_0}}\, n_i \Big/ \cos\theta_i, \quad \text{for parallel } (p) \text{ polarization,} \tag{3}$$

with the absolute dielectric constant and permeability of the free space ϵ_0 and μ_0, respectively. The whole layer system is described by the product of the N individual characteristic matrices taken in the correct order:

$$\mathcal{M} = \prod_{i=1}^{N} \mathcal{M}_i = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}. \tag{4}$$

The complex amplitude coefficient of reflection is then given by the four components $m_{11},\ldots,m_{22}$ of $\mathcal{M}$ and the optical admittance of the fiber substrate γ_0 and of the surrounding water γ_{N+1} ($n_0=1.48$, $n_{N+1}=1.329$):

$$r = \frac{\gamma_0 m_{11} + \gamma_0\gamma_{N+1} m_{12} - m_{21} - \gamma_{N+1} m_{22}}{\gamma_0 m_{11} + \gamma_0\gamma_{N+1} m_{12} + m_{21} + \gamma_{N+1} m_{22}}. \tag{5}$$

The intensity reflectance is $R = r \cdot r^*$, the asterisk denoting a complex conjugate.

In Fig. 2 the calculated reflectance R is depicted in dependence on the normalized design wavelength λ_D/λ for a 19-layer system with $n=2.30$ for the high index (Nb_2O_5) and $n=1.48$ for the low index material (SiO_2), for three different angles of light incidence in air $\theta_{in}=0°$, 25°, and 35°. For normal incidence of light, the high-finesse interference filter design produces a narrow transmission peak at $\lambda = \lambda_D = 2n_{10}d_{10} = 4n_i d_i$ ($i=1,...,9,11,...,19$). An incident pressure wave causes modulation of the optical thicknesses

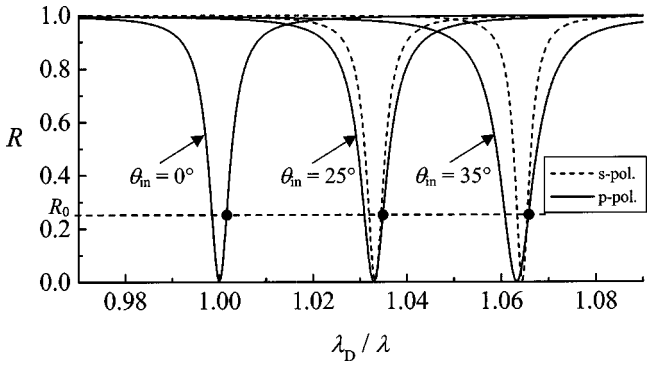


FIG. 2. Optical reflectance of the 19-layer system in dependence on the normalized optical thickness λ_D/λ for three different angles of light incidence θ_{in} for parallel (p) and perpendicular (s) polarization. Operational points with maximum $\Delta R/p_{in}$ at the edge of the transmission peaks (●).

of the layers and, thus, a modulation of optical reflectance of the whole layer system. The operating point of maximum reflectance slope $\Delta R/p_{in}$ is located at the edge of the transmission peak, where the reflectance is about $R_0 \approx 0.25$.[1,5] An alteration of θ_{in} gives rise to a change of the light propagation angles θ_i in the substrate, the layers, and the water (cf. Fig. 1) to be calculated by an appropriate application of Snell's law. Due to the changes in the phase thicknesses ϕ_i of the layers [Eq. (1)], a shift of both the transmission peak and the operating point toward higher values of λ_D/λ is induced. This behavior offers an important advantage over a setup in which the light incidence is restricted to the perpendicular direction: By choice of the appropriate angle θ_{in} that depends on the laser wavelength λ and the design wavelength λ_D of the specific interference filter sample used, the microinterferometer can be precisely adjusted to the operating point for ultrasound detection without tuning of the laser wavelength. In general, an adjustment is necessary because thickness deviations of the central spacer layer that in the manufacturing process are not completely avoidable cause a shift of the design wavelength from the desired value.[1] In addition to the spectral shift, a variation of the width of the transmission peak of the interference filter in dependence on θ_{in} can be seen in Fig. 2, which is caused by the angle- and polarization-dependent optical admittances [Eq. (3)]. For p polarization the finesse decreases with increasing θ_{in} due to decreasing absolute values of the complex Fresnel reflectivity coefficients for each layer boundary, whereas for s polarization, the finesse increases due to increasing absolute values of the coefficients.

For a quantitative description of the reflectance slope $\Delta R/p_{in}$ it is assumed that a perpendicularly incident plane ultrasonic wave produces a pure longitudinal wave within the sensor with a displacement $w(z)$ in the layer system and the substrate in the axial direction only. Negligible radial displacement can be assumed on the axis of the acoustic field where the small laser spot is located. Because of the radial symmetry with respect to the z axis, no torsional component occurs. Since the layer thicknesses d_i $(i=1,...,N)$ in all practical cases are small with respect to the acoustic wavelengths Λ in the layer materials, the axial strain $\partial w(z)/\partial z$ can be assumed constant along the z axis within each layer. For realistic layer numbers N, the thickness of the whole layer system is also small compared with Λ in the system, and the axial stress can be assumed constant within the whole layer system and is given by the negative value of the pressure p in the water in front of the sensor. Because of the acoustic impedance mismatch at the surface of the sensor, a reflected acoustic wave occurs and the pressure p in front of the sensor is about 1.8 times the free-field pressure of the incident wave p_{in}.

The deformation leads to a change Δd_i in the physical thickness of the ith layer:

$$\Delta d_i = \left(\frac{\partial w}{\partial z}\right)_i d_i = -\frac{p}{\rho_i v_i^2} d_i, \tag{6}$$

where ρ_i and v_i denote the density and longitudinal wave velocity, respectively, of the ith layer. Since no exact material parameters for the sputtered layers were available, the bulk parameters are used for approximate calculation instead (SiO_2: $\rho = 2250\ \mathrm{kg\,m^{-3}}$, $v = 5800\ \mathrm{m\,s^{-1}}$; Nb_2O_5: $\rho = 4500\ \mathrm{kg\,m^{-3}}$, $v = 6220\ \mathrm{m\,s^{-1}}$).[5] However, a modulation of the refraction coefficient n_i of each layer and the substrate ($i = 0,...,N$) results from the strain according to the elasto-optic effect:[9]

$$\Delta n_i = -\frac{1}{2} n_i^3 p_{12i} \left(\frac{\partial w}{\partial z}\right)_i = \frac{1}{2} n_i^3 p_{12i} \frac{p}{\rho_i v_i^2}, \tag{7}$$

where p_{12i} denotes the relevant strain-optic coefficient of the ith medium. Here the value for optical quartz fibers $p_{12} = 0.27$ taken from Ref. 9 is used for both materials.[3] The effects of strain-induced birefringence are neglected. In addition to the effects of layer deformation, the modulation of the refractive index n_{N+1} in the water by the pressure p in front of the sensor contributes to a variation of the optical reflectance ΔR of the multilayer system:[10]

$$\Delta n_{N+1} = p \times 1.5 \times 10^{-4}\ \ \mathrm{MPa^{-1}}. \tag{8}$$

However, in the case of the 19-layer system, the latter contribution is very small compared with the interference effect in the layer system.

In Fig. 3(a) the maximum reflectance slope $D = \Delta R/p_{in}$ of the 19-layer system at the operational wavelength is depicted in dependence on the angle of light incidence θ_{in} calculated with Eqs. (1)–(8), where plane optical waves were assumed within the layer system. With increasing angle the maximum reflectance slope increases for s polarization and decreases for p polarization, depending on the variations of the optical finesse of the multilayer system described above. Thus, s polarization should preferably be used to achieve high sensitivity. Figure 3(b) shows the shift in the operational optical wavelength λ of maximum reflectance slope in dependence on the angle of light incidence. For a given design wavelength λ_D the operational wavelength λ decreases with increasing angle θ_{in}.

Since the thickness modulations are small compared with the optical layer thicknesses of the undeformed system, a linear relationship between pressure amplitude and change in optical reflectance is realized. The calculated change in reflectance ΔR in dependence on the positive peak pressure

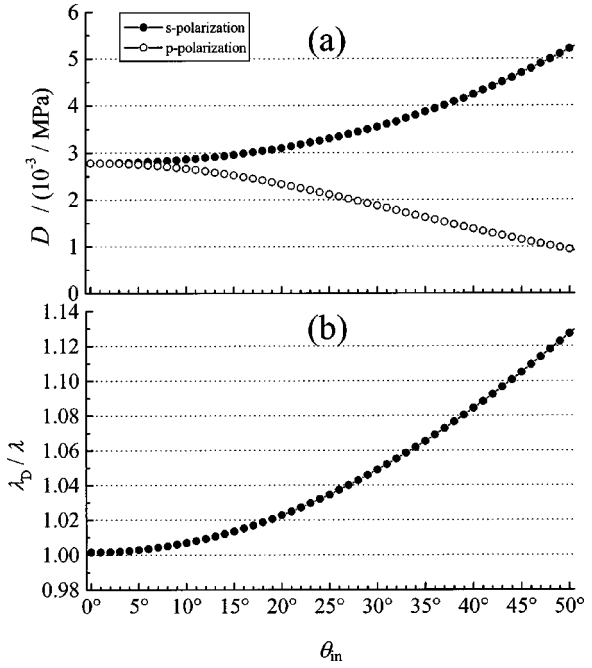


FIG. 3. Maximum reflectance slope $D=\Delta R/p_{\mathrm{in}}$ (a) of the 19-layer system at the appropriate operational wavelength (b) in dependence on the angle of light incidence θ_{in} for parallel (p) and perpendicular (s) polarization.

of the incident wave p_{in} for $\theta_{\mathrm{in}}=35°$ and s polarization using the same assumptions as before is shown in Fig. 4. Excellent linearity is found up to 20 MPa. A 5% decrease in D occurs at 37 MPa, and a 10% decrease at 53 MPa. Using different total layer numbers N, the linear range of the system can, in principle, be adjusted to the measurement task. It should be noted that the absolute values of D depend on the assumptions made for the material parameters that are expected to depend on the manufacturing process of the layers.

As mentioned above, all calculations were carried out for plane optical waves within the layer system, i.e., for single angles of light incidence. The results therefore are directly applicable to the use of a collimated laser beam in the optical detection scheme. However, to achieve high lateral resolution in the acoustic measurement, the laser needs to be focused. In this case the reflectance slope function obtained for a certain range of angles in the vicinity of θ_{in} and constant operational wavelength has to be weighted with the angle-dependent profile of the light intensity and integrated over this range of angles. The effect of this integration is an increase in R and a decrease in D with increasing divergence, which becomes significant at large θ_{in} and high optical finesse, in particular. In the practical realization the maximum reflectance slope achievable using large layer numbers N is, therefore, restricted by the oblique and focused illumination cone used.

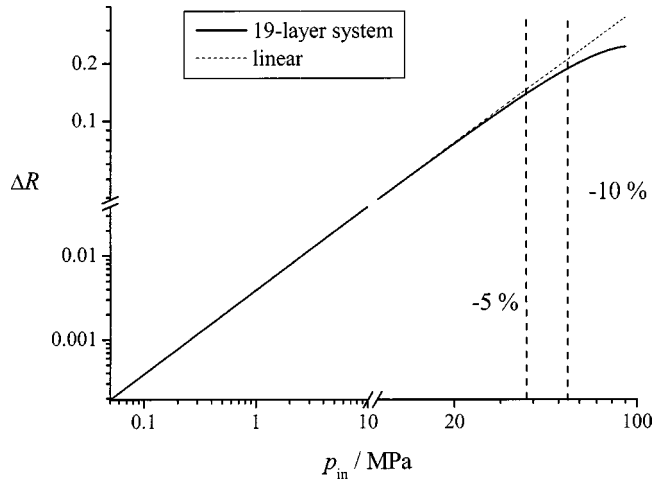


FIG. 4. The calculated change in reflectance ΔR of the 19-layer system in dependence on the acoustic positive peak pressure p_{in}; the range of linearity.

III. EXPERIMENTAL CHARACTERIZATION

Experiments were carried out to investigate the acoustic performance of the optical hydrophone. The multilayer system manufactured was made of $N=19$ layers with alternating refractive indices of $n=2.3\,(\mathrm{Nb_2O_5})$ and $n=1.48\,(\mathrm{SiO_2})$. The sputtering technique was employed to obtain a robust coating that can be used in water without swelling of the layers. The sensor element has an overall thickness of $d_{\mathrm{total}}\approx 1.9\,\mu$m. The design wavelength as measured by a grating spectrometer was ~675 nm, and a HeNe laser of 633 nm wavelength and ~2 mW output power was used in the optical detection setup. So, the angle of light incidence was adjusted to $\theta_{\mathrm{in}}\approx 35°$ [cf. Fig. 3(b)]. Perpendicular polarization of light was used and the focal length of the lens was $f_1=50$ mm (cf. Fig. 1). The measurement time is restricted to pulse or tone burst lengths up to ~2.2 μs, because after this transit time acoustic reflections from the rear side of the sensor substrate (thickness: 6.5 mm) may cause distortions. For longer measurement times a thicker substrate, a substrate with stronger acoustic absorption, or an additional, appropriately mounted absorber can be used.

A. Lateral resolution

The lateral resolution capability of the sensor system is mainly determined by the optical spot size of the focused light beam within the multilayer system. For an estimation of the spot dimension, a knife edge was fixed to the layer system, which was then moved in the horizontal direction perpendicular to the edge. A successive evaluation of the part of light transmitted in dependence on the edge position in the x and y directions showed a Gaussian intensity profile with full widths at $1/e$ of the maximum of ~60 μm in the x direction and ~45 μm in the y direction. The increased width in the x direction is first due to the oblique incidence of light, resulting in an elliptical cross section of the illuminating spot. A further enlargement of the width in the x direction is then brought about by the multi-interference process in the layer system.

It should be noted that the lateral distribution of the acousto-optic transduction efficiency within the focused optical spot depends not only on the Gaussian light intensity distribution but also on the optical angular spectrum, as described in Sec. II, and can be quite inhomogeneous in the x direction, in particular. Furthermore, in applications of ultrasound imaging using the sensor as a two-dimensional detection array, the lateral resolution may also be restricted by

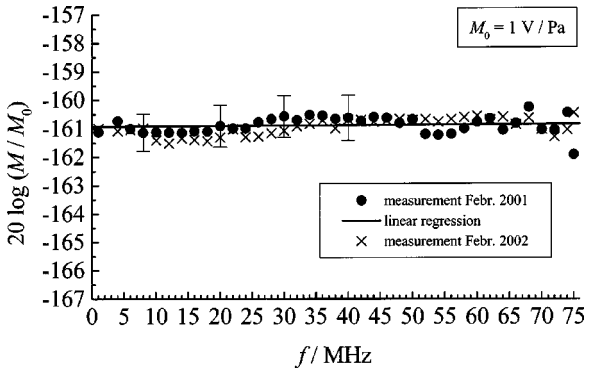


FIG. 5. The frequency response of the optical multilayer hydrophone as determined by the primary interferometric calibration; measurement repeated after one year; linear regression; uncertainty bars: standard uncertainties.

surface acoustic waves (cf. Sec. III D), a mechanical crosstalk phenomenon, which is known, for instance, from capacitive micromachined ultrasonic transducer arrays.

B. Frequency response

A technique for primary calibration of hydrophones based on a Michelson interferometer was applied to experimentally determine the frequency response of the optical multilayer hydrophone. As a detailed description of the calibration setup and the evaluation procedure is given in Ref. 11, a brief description will be sufficient here. A stabilized interferometer detected the displacement of a pellicle (thickness: 2 μm) mounted on the surface of the water contained in a tank, while ultrasound bursts from a focusing PVDF source transducer (diameter: 5 mm, focal length: 12.5 mm, center frequency: 50 MHz) were incident from the bottom. The pressure amplitudes p_{in} were calculated from the displacement amplitudes measured in the focal plane for discrete frequencies in the range from 1 to 75 MHz. In a second course, the known acoustic field at each frequency was measured using the optical multilayer hydrophone and the same excitation conditions of the source transducer as in the interferometric measurements. The transfer function $M(f) = U/p_{\mathrm{in}}$ was then calculated by a comparison of the signal voltage amplitudes U with the reference measurement results. A data correction was carried out to compensate the acoustic transmission characteristic of the foil and the frequency response of the photodetector in the interferometric reference setup. In addition, a correction was performed to compensate the frequency response of the photodetector (including amplifier) of the multilayer hydrophone determined in a separate optical mixing measurement.[11] Since the optical spot size of the two detection systems was approximately the same, corrections for the influence of spatial averaging of the focused sound field could be omitted.

The pressure–voltage transfer function obtained is shown in Fig. 5. The measurement uncertainties (66% confidence level) depicted for several frequencies are derived from the uncertainties of the voltage measurements (including the signal-to-noise ratio) and sensor positioning, from an estimation of sound field inhomogeneities, and from the uncertainties of the different correction procedures referred to above.[11] Within the uncertainty of the calibration, a constant transfer function of $M(f) = 9.0\,\mathrm{mV/MPa}$ is obtained for the whole frequency range investigated. Due to the small thickness of the sensing element ($d_{\mathrm{total}} \approx 1.9\,\mu\mathrm{m}$), no thickness resonances occur (subresonant mode) that are known to bring about a strong variation of the transfer function, for instance, in the case of membrane hydrophones with a typical foil thickness of 25 μm in the frequency range of about 40 MHz.[11] Thickness mode resonance distortions are also known from other optical ultrasound detectors like low-finesse polymer film Fabry–Perot-type[12,13] and high-finesse etalon-type sensors.[6] Although the acoustic frequency response has not been investigated as a function of the etalon thickness in Ref. 6, resonances can be expected according to the acoustic sensitivity of both the front and the rear face of the sensor element, which can be observed in the impulse measurements performed. Since the diameter of the optical multilayer hydrophone sensor probe is much larger than that of the sound field, edge diffraction, lateral waves, and radial resonances within the sensor element do not arise, whereas these effects are known to cause distortions of the frequency response of small probe hydrophones like, for instance, needle-type hydrophones[14,15] and, in particular, fiber-optic multilayer hydrophones.[3]

It is appropriate to compare the experimentally obtained transfer factor with the theoretical result according to the analysis in Sec. II. The maximum reflectance slope at $\theta_{\mathrm{in}} = 35°$ for s polarization was determined to be $D = 3.87 \times 10^{-3}/\mathrm{MPa}$ [cf. Fig. 3(a)]. The diameter of the Gaussian profile laser beam incident on the focusing lens ($f = 50\,\mathrm{mm}$), measured by a procedure similar to that described in Sec. III A, was $d = 550\,\mu\mathrm{m}$ (full width at $1/e$ of maximum). Numerical integration of the reflectance function and the reflectance slope function, respectively, weighted with this geometrical illumination cone incident at $\theta_{\mathrm{in}} = 35°$ yields $R_{\mathrm{cone}} = 1.06 \times R_0$ and $D_{\mathrm{cone}} = 0.72 \times D$. Using $D_{\mathrm{cone}} = \Delta R_{\mathrm{cone}}/p_{\mathrm{in}}$ together with $\Delta R_{\mathrm{cone}}/R_{\mathrm{cone}} = 2.8 \times U/U_0$, where $U_0 = 2.35\,\mathrm{V}$ is the direct current (dc) voltage of the photodetector corresponding to the reflectance R_{cone} at the operating point and 2.8 the ratio of electrical amplification of the dc output to the alternating current (ac) output of the photodetector, the theoretical pressure–voltage transfer factor is given by

$$M_{\mathrm{th}} = \frac{U}{p_{\mathrm{in}}} = \frac{U D_{\mathrm{cone}}}{\Delta R_{\mathrm{cone}}} = \frac{D_{\mathrm{cone}} U_0}{2.8 R_{\mathrm{cone}}} = 8.8\ \ \mathrm{mV/MPa}, \tag{9}$$

which is in good agreement with the experimental value.

Note that the absolute values given in Fig. 5 on the ordinate depend on the amplification chosen for the photocurrent and in this sense are arbitrary and cannot be compared with values known from PVDF hydrophones to estimate the sensitivity. This comparison is dealt with in the following paragraph.

C. Sensitivity

The sensitivity of the optical multilayer hydrophone system was determined by measurement of the noise voltage U_{noise} of the photodetector output. From the noise voltage, as

TABLE I. Sensitivity in terms of the noise-equivalent peak pressure values measured for the present optical multilayer hydrophone system, a PVDF membrane hydrophone/amplifier combination, and a small diameter needle-type PVDF hydrophone.

Measurement system	Noise-equivalent peak pressure p_{min} (25 MHz detection bandwidth; without averaging)
Optical multilayer hydrophone	30 kPa
0.5 mm diameter PVDF membrane hydrophone	0.6 kPa
0.075 mm diameter PVDF needle-type hydrophone (Ref. 13)	55 kPa

measured by an oscilloscope, the noise-equivalent pressure was obtained using the transfer factor from the experimental calibration: $p_{min}=U_{noise}/M$. The noise-equivalent acoustic pressure of the multilayer hydrophone for an analog detection bandwidth of the oscilloscope of ~25 MHz (sampling rate 250 MS/s) without signal averaging is given in Table I, in comparison with the corresponding value of a typical 0.5 mm diameter PVDF membrane hydrophone/amplifier system. In addition, the value taken from Ref. 13 for a 75 μm diameter PVDF needle-type hydrophone obtained in a similar way is listed. It should be noted that the noise sources of the different measurement systems are of a different kind. While the noise during the PVDF hydrophones measurement is dominated by that from the amplifier used (separately or within the oscilloscope), the noise during the optical measurement is mainly made up of light intensity noise from the laser source and shot noise from the optical detection process. The results show that on the one hand the sensitivity of the present optical multilayer hydrophone is significantly lower (1.5 orders of magnitude) than that of the membrane hydrophone that is 0.5 mm in diameter. On the other hand, the PVDF needle-type hydrophone with a sensor diameter comparable to that of the optical hydrophone shows a sensitivity smaller by approximately a factor of 2. In general, the sensitivity of smaller diameter PVDF hydrophones is lower due to the lower capacitance. In principle, this trade-off between sensitivity and lateral resolution can be circumvented when using the optical multilayer hydrophone. Here the sensitivity depends on the reflectance slope of the layer system and on the power and noise characteristics of the light source but not on the size of the sensing element. This applies similarly to sensitivity versus frequency response: To achieve a large frequency range with a flat transfer function, it is necessary to use small foil thicknesses for the PVDF hydrophones (subresonant mode), which, in turn, produce low sensitivity. This second trade-off can also be circumvented when using the optical hydrophone that inherently uses a very small overall thickness, even for large layer numbers.

D. Directional response

To determine the directional response of the multilayer hydrophone a focusing source transducer (diameter: 6 mm; focal length: 20 mm) was turned about the measurement point at the water surface, since the present detection system does not allow the detector to be turned. For three different frequencies (10, 15, 25 MHz) tone bursts were measured in angular steps of 2° and the voltage amplitudes were related to the amplitude at normal sound incidence. Measurements were successively made for turning in the xz plane and in the yz plane (cf. Fig. 1). The results obtained are shown in Fig. 6. Obviously, the directional response does not significantly depend on the ultrasound frequency in the frequency and angular ranges investigated. This is in contrast to the directional response known, for instance, from membrane hydrophones, which is, with the exception of the impact of Lamb waves at large angles (~55°) and low frequencies, well described by the diffraction theory of a plane rigid circular disk receiver.[16] The directional response of the optical multilayer hydrophone is not dominated by diffraction but likely by the impact of Rayleigh waves excited on the sensor surface. In the range from 20° to 30°, the excitation is most effective because here the propagation constant components parallel to the sensor surface of the incident wave and of the surface wave are matched, which results in a strong increase of the response in this range. The asymmetric profile found for the xz plane is a result of the asymmetric variation of the reflectance slope in dependence on the angle of illumination for the optical cone at oblique incidence (cf. Sec. II). In contrast, the variation of the reflectance slope within the cone of illumination in the perpendicular direction is very small and symmetric to the z axis, leading to a symmetric directional response in the yz plane.

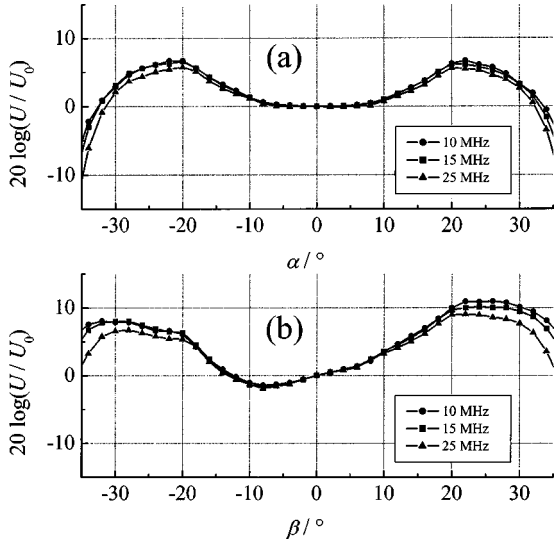


FIG. 6. Directional response of the optical multilayer hydrophone measured perpendicular (yz plane, α), (a) and parallel (xz plane, β), (b) to the plane of light incidence (cf. Fig. 1) for different ultrasound frequencies; U: voltage amplitude, U_0: voltage amplitude at normal incidence.

E. Stability

The short-term stability was investigated by repeated measurements of ultrasound pulses from a focusing source transducer (center frequency: 8 MHz, diameter 6 mm, focal length: 15 mm) driven by a pulse generator (peak voltage: 350 V). After 60 min of warming up for the laser source, the sensor, which was at the same temperature as the water, was positioned at the water surface and the peak voltage of a

20-times-averaged waveform was immediately acquired. The measurement was randomly repeated over several hours. The maximum relative deviation of the peak voltage was 2% and a systematic tendency with time could not be identified. The deviation of the laser power simultaneously measured was in the range of 1%. With respect to other statistical deviations caused, for instance, by the source transducer and the pulse generator used, better stability could not be expected. Significant swelling of the layers that would have changed the optical reflectance was not observed.

The long-term stability of the optical multilayer hydrophone was confirmed when the interferometric calibration was repeated after one year, and so far no significant deviation due to degradation of the sensor could be observed (Fig. 5).

IV. CONCLUSIONS

An optical multilayer hydrophone for broadband ultrasound detection was theoretically described and experimentally characterized. The acoustic measurement is based on the elastic deformation of the very thin optical layer system deposited onto a glass substrate by an incident pressure wave and the detection of the induced change in optical reflectance. Using an oblique incidence of light to detect the reflectance variation yields good sensitivity and a very simple and low-cost setup. For the theoretical description some simplifying assumptions and material parameter estimations were made to study the sensor performance and the opportunity to adjust the operational optical wavelength of a given layer system by variation of the angle of light incidence. From this several design criteria for the detection arrangement could be derived.

A concrete optical multilayer hydrophone arrangement was experimentally characterized. The lateral resolution, the frequency response, the sensitivity, the directional response, and the stability were determined. Because of the very small sensor element thickness and the absence of edge and radial effects, the frequency response obtained is constant in a remarkably broad frequency range, which was confirmed for the range from 1 to 75 MHz. A comparison of the experimental pressure–voltage transfer factor with the theoretical estimation showed very good agreement. The directional response is dominated by the impact of Rayleigh waves that are excited when the sound wave obliquely strikes the sensor surface. The sensitivity of the present system is comparable to that of available PVDF hydrophones with similar lateral dimensions of the sensor element.

The results, in particular, the constant response in a wide frequency range, show that the optical multilayer hydrophone meets very well the demands placed on a reference measurement system for secondary calibrations. By measurement of single broadband ultrasound pulses (wave fields rich in harmonics) with the optical hydrophone and in a second step with a measurement system to be calibrated, the frequency response can be determined by a comparison of the relevant frequency spectra of the measured pulses. The complex-valued frequency response is obtained with high-frequency resolution in a very fast and efficient way and can be used for correct data evaluation and impulse deconvolution. This technique has already been successfully applied to the calibration of fiber-optic multilayer hydrophones using a focused source transducer with a center frequency of 8 MHz, and evaluation up to the fifth harmonic provided the frequency response up to 40 MHz.[5,17] The technique may be extended also to the calibration of PVDF hydrophones using broadband pulses.[18–21] Furthermore, a direct comparison with the constant frequency response of the optical multilayer hydrophone would eliminate the need for extended sound field calculations, including idealized assumptions for the source transducer, in broadband calibration techniques without a reference receiver using nonlinear sound propagation.[22] The optical multilayer hydrophone may also be regarded as a possible independent method to confirm the existing interferometric calibration techniques, in particular, as regards their extension to the high-frequency range recently developed.[11,23] From the metrological perspective, the comparison of piezoelectric hydrophones with the acoustic pressure detecting optical hydrophone is more advantageous than a comparison with interferometers detecting particle displacement, since for a given acoustic pressure, the particle displacement for a plane wave is reciprocally related to the frequency, which imposes more severe requirements for an acceptable signal-to-noise ratio at high frequencies.[23]

Another application field for the detection system presented may lie in high-frequency ultrasound imaging, where detection arrays with a small element size are called for.[2,24,25] A promising concept in this field of nondestructive testing is the application of broadband optical ultrasound detection systems in combination with laser-induced photoacoustic pulse generation.[7,26,27]

ACKNOWLEDGMENT

The author wishes to thank Walter Molkenstruck for support with regard to the interferometric calibration.

[1] V. Wilkens and C. Koch, ''Fiber-optic multilayer hydrophone for ultrasonic measurement,'' Ultrasonics **37**, 45–49 (1999).
[2] V. Wilkens and C. Koch, ''Optical multilayer detection array for fast ultrasonic field mapping,'' Opt. Lett. **24**, 1026–1028 (1999).
[3] W. Weise, V. Wilkens, and C. Koch, ''Frequency response of fiber-optic multilayer hydrophones: experimental investigation and finite element simulation,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 937–946 (2002).
[4] V. Wilkens and C. Koch, ''Optische Vielschichtsensoren: eine Alternative zu piezoelektrischen Ultraschallhydrophonen,'' *Fortschritte der Akustik—DAGA 2000*, Proceedings of the 26th German Annual Conference on Acoustics, Oldenburg, Germany, ISBN 3-9804568-8-9, 680-681, 2000 (in German).
[5] V. Wilkens, ''Dielektrische optische Interferenzschichthydrophone zur zeitlich und räumlich hochauflösenden Messung von Ultraschallfeldern,'' Ph.D. dissertation, University of Oldenburg, Germany, 2001, and PTB Report No. PTB-MA-67, 2001, ISBN 3-89701-750-4 (in German).
[6] J. D. Hamilton, T. Buma, M. Spisar, and M. O'Donnell, ''High frequency optoacoustic arrays using etalon detection,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 160–169 (2000).
[7] P. C. Beard, ''Photoacoustic imaging of blood vessel equivalent phantoms,'' *Proceedings of the SPIE*, BIOS 02, San Jose, CA, 2002, Vol. 4618, pp. 54–62.
[8] H. A. Macleod, *Thin-Film Optical Filters* (Institute of Physics, Bristol, 2001), Chap. 2, pp. 37–45.
[9] G. B. Hocker, ''Fiber-optic sensing of pressure and temperature,'' Appl. Opt. **18**, 1445–1448 (1979).
[10] J. Staudenraus and W. Eisenmenger, ''Fibre-optic probe hydrophone for

ultrasonic and shock-wave measurements in water,'' Ultrasonics **31**, 267–273 (1993).
[11] C. Koch and W. Molkenstruck, ''Primary calibration of hydrophones with extended frequency range 1 to 70 MHz using optical interferometry,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 1303–1314 (1999).
[12] P. C. Beard, F. Pérennès, and T. N. Mills, ''Transduction mechanisms of the Fabry–Perot polymer film sensing concept for wideband ultrasound detection,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 1575–1582 (1999).
[13] P. C. Beard, A. M. Hurrell, and T. N. Mills, ''Characterization of a polymer film optical fiber hydrophone for use in the range 1 to 20 MHz: a comparison with PVDF needle and membrane hydrophones,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 256–264 (2000).
[14] B. Fay, G. Ludwig, C. Lankjaer, and P. A. Lewin, ''Frequency response of PVDF needle-type hydrophones,'' Ultrasound Med. Biol. **20**, 361–366 (1994).
[15] Y. Nakamura and T. Otani, ''Study of surface elastic wave induced on backing material and diffracted field of a piezoelectric polymer film hydrophone,'' J. Acoust. Soc. Am. **94**, 1191–1199 (1993).
[16] S. Robinson, R. Preston, M. Smith, and C. Millar, ''PVDF reference hydrophone development in the UK—from fabrication and lamination to use as secondary standards,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1336–1344 (2000).
[17] V. Wilkens, C. Koch, and W. Molkenstruck, ''Frequency response of a fiber-optic dielectric multilayer hydrophone,'' *Proceedings of the 2000 IEEE Ultrasonics Symposium*, San Juan, Puerto Rico, 2000, pp. 1113–1116.
[18] R. Reibold and W. Molkenstruck, ''Investigation of pulse-excited hydrophones for ultrasonic field measurements using laser interferometry,'' Ultrasonics **25**, 114–118 (1987).
[19] H. Djelouah, J. C. Baboux, and M. Perdrix, ''Pulsed calibration technique of miniature ultrasonic receivers using a wideband laser interferometer,'' Ultrasonics **27**, 80–85 (1989).
[20] G. R. Harris and P. M. Gammell, ''Sensitivity measurements of piezoelectric polymer hydrophones from 0.2 to 2 MHz using a broadband pulse technique,'' J. Acoust. Soc. Am. **105**, 725–731 (1999).
[21] R. A. Smith and D. R. Bacon, ''A multiple-frequency hydrophone calibration technique,'' J. Acoust. Soc. Am. **87**, 2231–2243 (1990).
[22] H. J. Bleeker and P. A. Lewin, ''A novel method for determining calibration and behaviour of PVDF ultrasonic hydrophone probes in the frequency range up to 100 MHz,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1354–1362 (2000).
[23] T. J. Esward and S. P. Robinson, ''Extending the frequency range of the NPL primary standard laser interferometer for hydrophone calibration to 60 MHz,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 737–744 (1999).
[24] J. D. Hamilton and M. O'Donnell, ''High frequency ultrasound imaging with optical arrays,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 216–235 (1998).
[25] P. C. Beard and T. N. Mills, ''A 2D optical ultrasound array using a polymer film sensing interferometer,'' in Ref. 17, pp. 1183–1186.
[26] G. Paltauf and H. Schmidt-Kloiber, ''Measurement of laser-induced acoustic waves with a calibrated optical transducer,'' J. Appl. Phys. **82**, 1525–1531 (1997).
[27] K. P. Köstli, M. Frenz, H. P. Weber, G. Paltauf, and H. Schmidt-Kloiber, ''Optoacoustic tomography: time-gated measurement of pressure distributions and image reconstruction,'' Appl. Opt. **40**, 3800–3809 (2001).

New modeling method and mechanism analyses for active control of interior noise in an irregular enclosure using piezoelectric actuators

Hou C. Geng
State Key Laboratory of Vibration, Shock & Noise, Shanghai Jiao Tong University, 1954 Hua Shan Road, Shanghai 200030, People's Republic of China and Postdoctoral R&D Center, Jiangnan Shipyard Co. Ltd., 2 Gao Xiong Road, Shanghai 200011, People's Republic of China

Zhu S. Rao and Zu S. Han
State Key Laboratory of Vibration, Shock & Noise, Shanghai Jiao Tong University, 1954 Hua Shan Road, Shanghai 200030, People's Republic of China

(Received 28 October 2002; accepted for publication 12 November 2002)

A new modeling method is developed in this paper for the active minimization of noise within a three-dimensional irregular enclosure using distributed lead zirconate titanate piezoelectric (PZT) actuators, and the control mechanisms for irregular enclosure are analyzed. The irregular enclosure is modeled with four rigid walls and two simply supported flexible panels, and PZT actuators are bound to one of the flexible panels. The process of the new modeling method is as follows. First, the modal coupling method is used to establish the motion equations, which contain important coefficients such as modal masses and modal coupling coefficients, etc., of acoustic–structural–piezoelectric coupling system. Then, the acoustic modes and the modal masses of irregular enclosure are calculated by numerical methods. Last, the modal coupling coefficients in motion equations are calculated according to the numerical results of the acoustic modes of irregular enclosure and the modes of two panels. The validity of this modeling method is verified by a regular hexahedron enclosure. Two cost functions are applied to this model. With the two cost functions, good results are obtained in minimizing the sound-pressure level (SPL) within irregular enclosure according to numerical investigations. By comparing the results obtained under controlled and uncontrolled states, the control mechanisms of the system are discussed. It is found that the control mechanisms vary with disturbance frequencies. At most disturbance frequencies, the SPL within enclosure is reduced by restructuring the modes of two panels simultaneously. When the disturbance frequency comes close to one of the natural frequencies of panel *a*, the dominant mode of panel *a* is suppressed, while the modes of panel *b* are reconstructed. While the disturbance frequency is near one of the natural frequencies of panel *b*, the modes of two panels are restructured at the same time.
 [DOI: 10.1121/1.1543585]

PACS numbers: 43.40.At [PJR]

I. INTRODUCTION

An important problem in systems such as automobile, aircraft, and rotorcraft, etc. is the control of interior noise with low frequencies, which is generated by the vibration of elastic structures. Passive noise-control approaches such as sound absorption and damping treatment, etc. have been deemed as the main attenuation methods in the past decades. However, the effectiveness of passive noise control is limited due to the variations of damping material properties with temperature and frequencies, and the sound energy in low frequencies is difficult for absorption because the wavelength of sound is relatively large compared with the thickness of absorbing materials. Hence, the active noise-control technology, which is proven to be more effective in low frequencies, has been gaining rapid development in recent years.[1]

In general, the active noise-control approach involves two branches. In one branch, a number of secondary acoustic sources are arranged around primary noise source to cancel out the radiated sound.[2–5] In another branch, a number of actuators are applied directly to structures themselves.[6,7] Piezoelectric materials are used as actuators and/or sensors[8–15] in most cases.

Both the analytical modeling and numerical modeling method are used in minimization of noise within enclosures. A cubic cavity model with five rigid walls and one flexible thin plate cover is most frequently discussed. Piezoelectric materials are bound to the plate cover to control structure-borne radiations. The analytical modeling method is based on classic wave motion equation and elastic structure vibration theory. Using this method and the above model, Sampath *et al.*,[8] Banks *et al.*,[9] and Koshigoe[10,11] investigated the transmission of noise from an external source into cavity and their controls. Green's functions were used to establish the equations for coupling plate–acoustic interaction. A similar work but with a point primary source inside the cavity has also been presented by Bao *et al.*,[12] in which the reduction of SPL is 15–22 dB. The numerical modeling method was discussed in the works of Poh *et al.*,[13] Shields *et al.*,[14] and Kawamura *et al.*[15] In these studies, motion equations were established by finite element and boundary element methods, and the actuators and/or sensors made of smart materials

were described by discrete elements. In general, the numerical modeling method can be used to deal with an enclosure with rather complicated boundaries. However, it is difficult to be used in real time control. The analytical modeling method is restricted to application in regular hexahedron enclosures, because up to the present only in these regular enclosures can acoustic modes be described with theoretical expressions. As mentioned above, the two modeling methods do have their inherent defects. Finding a new modeling method for irregular enclosures is a challenging task.

In this paper, a new modeling method for an irregular enclosure is proposed. An irregular enclosure with four rigid walls and two flexible panels is considered, in which several point forces are applied on one of the panels and a number of piezoelectric actuators are bound to the other. In system modeling, modal coupling method is applied to derive motion equations, provided that the acoustic modes of irregular enclosure exist. The modal coupling method has been proven to be one of the well-established analytical methods for accurately evaluating the response of acoustic–structural coupling system.[16,17] Numerical method is used to calculate the acoustic modes and acoustic modal masses of irregular enclosure. Modal coupling coefficients are crucial for this new modeling method, and they are calculated based on the numerical results of the acoustic modes of irregular enclosure and the analytical mode shapes of two simply supported panels. That is to say, the values of the modal coupling coefficients and the acoustic modal masses in motion equations are substituted by their relative numerical results. The motion equations established in this paper can be used to investigate the characteristics of the active noise control equally as that of regular enclosures, and is not subject to the limitation of enclosure boundaries.

In addition to modeling studies there has been a large number of papers analyzing the control mechanisms of enclosures. Through analytical and experimental studies, Pan *et al.*[18,19] obtained two different control mechanisms. If the system response is dominated by panel modes, the sound energy in the cavity is minimized by suppressing the panel modes (modal suppression). If the system response is dominated by cavity modes, the control force is used to change the panel velocity distribution and thus adjust the radiation of each panel mode (modal restructuring). Qiu *et al.*[20] discussed the same model further, and concluded that these two control mechanisms usually coexist. It is further pointed out that when the cost function of the minimization is chosen as $\langle pp^* \rangle$, in which p is the sound pressure within cavity, and one point force at the panel center is used as active control source, the control mechanism is different from what is given in Refs. 18 and 19 at some resonance frequencies.

In this paper, the control mechanisms for irregular enclosure are also elucidated when a disturbance point force is applied at the center of panel *b* and one pair of PZT patches is bound to panel *a*. Results obtained demonstrate that for this irregular enclosure, the control mechanisms are more complex than that given in Ref. 18. At most disturbance frequencies, the SPL within the enclosure is decreased by restructuring the modes of two panels simultaneously. When the control frequency is near one of the natural frequencies of panel *a*, the SPL within the enclosure is reduced by suppressing the dominant mode of panel *a* and restructuring the modes of panel *b*, respectively. When the control frequency is near one of the natural frequencies of panel *b*, the SPL within enclosure is reduced by restructuring the modes of two panels simultaneously.

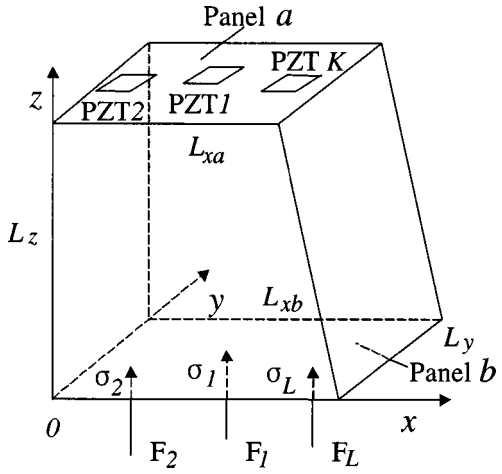


FIG. 1. Geometry and coordinate system of an irregular enclosure with two elastic panels.

II. SYSTEM MODELING

Figure 1 shows the irregular enclosure used in this paper. The irregular enclosure has four rigid boundaries and two simply supported elastic panels. The two panels, labeled as panel *a*, panel *b*, and located at $z=L_z$ and $z=0$, respectively, form the flexible boundaries of the irregular enclosure. These panels are assumed to be thin and isotropic. *K* pairs of PZT patches are bound symmetrically to the top and the bottom of panel *a* to form actuators, and the patches are assumed to have the same geometric and material characteristics. *L* external disturbing point forces are applied vertically on the bottom of panel *b*.

A. In enclosure

The inhomogeneous wave equation inside the enclosure is given by

$$\nabla^2 p - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} = -\rho_0 \frac{\partial q}{\partial t}, \tag{1}$$

where p is the sound pressure within enclosure, ∇^2 is the Laplacian operator, ρ_0 and c_0 are the equilibrium fluid density and the sound velocity within the enclosure, q is the distribution of sound source volume velocity per unit volume. The vibrations of panel surfaces may be represented as two volume velocity distributions confined to related infinitesimally thin layers situated on surfaces ξ_a and ξ_b just inside two rigid wall boundaries. Assume that normal structural displacement is positive outward, then we have

$$q = -\frac{\partial w_a}{\partial t}\delta(\xi - \xi_a) - \frac{\partial w_b}{\partial t}\delta(\xi - \xi_b), \tag{2}$$

where w_a and w_b are the displacements of panels *a* and *b*, respectively, $\delta(\cdot)$ is the one-dimensional Dirac function in

the direction normal to surface. Expanding p in terms of the time-dependent acoustic modal coordinates p_n and the acoustic mode shapes ψ_n for the fluid volume with rigid boundaries, we have

$$p=\sum_n p_n\psi_n, \tag{3}$$

with the following properties

$$\nabla^2\psi_n+(\omega_n/c_0)^2\psi_n=0, \tag{4a}$$

$$\int_V \psi_i\psi_n\, dV=\begin{cases}0, & i\neq n\\ M_n, & =n,\end{cases} \tag{4b}$$

where ω_n is the nth acoustic natural frequency, V represents the volume of enclosure. Using Eqs. (2), (3), and (4) in Eq. (1), and integrating over the fluid volume, we obtain

$$\ddot{p}_n+\omega_n^2 p_n=\frac{\rho_0 c_0^2}{M_n}\int_{A_a}\psi_n\left[\frac{\partial^2 w_a}{\partial t^2}\right]dA - \frac{\rho_0 c_0^2}{M_n}\int_{A_b}\psi_n\left[\frac{\partial^2 w_b}{\partial t^2}\right]dA, \tag{5}$$

where A_a and A_b refer to the areas of surfaces a and b, respectively.

Expanding the panels' displacements in terms of their modal coordinates q_r^a, q_s^b, and *in vacuo* normal modes φ_r^a, φ_s^b as

$$w_a=\sum_r q_r^a\varphi_r^a, \quad w_b=\sum_s q_s^b\varphi_s^b, \tag{6}$$

with the properties

$$\begin{aligned}&D_a\nabla^4(\varphi_r^a)-\omega_r^{a^2}m_a\varphi_r^a=0\\ &D_b\nabla^4(\varphi_s^b)-\omega_s^{b^2}m_b\varphi_s^b=0,\end{aligned} \tag{7a}$$

$$\int_{A_a} m_a\varphi_i^a\varphi_r^a\, dA=\begin{cases}0, & i\neq r\\ M_r^a, & i=r,\end{cases} \tag{7b.1}$$

$$\int_{A_b} m_b\varphi_i^b\varphi_s^b\, dA=\begin{cases}0, & i\neq s\\ M_s^b, & i=s,\end{cases} \tag{7b.2}$$

in which r refers to the rth mode of panel a and s the sth of panel b, ω_r^a and ω_s^b are the rth and sth *in vacuo* natural frequencies of panel a and panel b, respectively. $\nabla^4=\nabla^2\nabla^2$, $D_a=E_a h_a^3/12(1-\nu_a)$, $D_b=E_b h_b^3/12(1-\nu_b)$; m_a, h_a, E_a, ν_a and m_b, h_b, E_b, ν_b represent the mass per unit area, thickness, Young's modulus, and Poisson's ratio of panels a and b, respectively. Substituting Eqs. (6) and (7) into Eq. (5), we can get the motion equations for fluid in enclosure

$$\ddot{p}_n+\omega_n^2 p_n=-\frac{\rho_0 c_0^2 A_a}{M_n}\sum_r C_{nr}^a\ddot{q}_r^a-\frac{\rho_0 c_0^2 A_b}{M_n}\sum_s C_{ns}^b\ddot{q}_s^b, \tag{8}$$

where C_{nr}^a and C_{ns}^b are dimensionless modal coupling coefficients

$$C_{nr}^a=\frac{1}{A_a}\int_{A_a}\varphi_n\varphi_r^a dA, \quad C_{ns}^b=\frac{1}{A_b}\int_{A_b}\varphi_n\phi_s^b dA. \tag{9}$$

B. For panel *a*

The motion equation of panel a, excited by PZT patches with input control voltages $V_k(t)$ can be expressed as[21]

$$\begin{aligned}&\rho_e h_e\ddot{w}_a+D_a\nabla^4(w_a)+D_a\alpha R_E\times\sum_k\left\{\frac{\partial^2}{\partial x^2}\left[\chi_k\left(\frac{\partial^2 w_a}{\partial x^2}+\nu_a\frac{\partial^2 w_a}{\partial y^2}\right)\right]+\frac{\partial^2}{\partial y^2}\left[\chi_k\left(\nu_a\frac{\partial^2 w_a}{\partial x^2}+\frac{\partial^2 w_a}{\partial y^2}\right)\right]\right.\\ &\left.+2(1-\nu_a)\frac{\partial^2}{\partial x\partial y}\left(\chi_k\frac{\partial^2 w_a}{\partial x\partial y}\right)\right\}\\ &=p(\xi_a,t)-\sum_k\frac{(h_a+h_p)E_p d_{31}}{(1-\nu_p)}(\nabla^2\chi_k)V_k(t),\end{aligned} \tag{10}$$

where $\rho_e h_e=\rho_a h_a(1+2\Sigma_{k=1}^K\rho_p\xi_h\chi_k/\rho_a)$, $\xi_h=h_p/h_a$, $\alpha=6\xi_h+12\xi_h^2+8\xi_h^3$, $R_E=E_p/E_a$. The subscript p stands for PZT patches and subscript a for panel a; ρ_p, h_p, E_p, ν_p, and d_{31} represent the density, thickness, Young's modulus, Poisson's ratio, and dielectric constant of PZT patches, respectively. The function χ_k is unity, where the kth PZT patches exist and is zero otherwise, and K is the number of PZT pairs. The term $p(\xi_a,t)$ is the acoustic pressure on the low surface of panel a, and $V_k(t)$ is the control voltage applied to the kth PZT pair. Now, we assumed that the thickness of the PZT patches is negligible compared with that of panels, the density of two materials is approximately the same, and the size of the actuators is small compared with the wavelengths considered. Therefore, the stiffness term on the left-hand side of Eq. (10) can be ignored, and $\rho_e h_e$ is replaced by $\rho_a h_a(=m_a)$. Hence, we obtain

$$\begin{aligned}&m_a\ddot{w}_a+D_a\nabla^4(w_a)\\ &=p(\xi_a,t)-\sum_k\frac{(h_a+h_p)E_p d_{31}}{(1-\nu_p)}(\nabla^2\chi_k)V_k(t).\end{aligned} \tag{11}$$

Using Eqs. (3), (6), and (7) in Eq. (11), followed by multiplication by φ_r^a and integration over the surface of panel a, we get the coupling motion equations of panel a

$$\ddot{q}_r^a+\omega_r^{a^2}q_r^a=\frac{A_a}{M_r^a}\sum_n p_n C_{nr}^a-\frac{1}{M_r^a}\sum_k\epsilon_{rk}V_k, \tag{12}$$

where

$$\epsilon_{rk}=\frac{(h_a+h_p)E_p d_{31}}{(1-\nu_p)}\int_{A_a}\varphi_r^a(\nabla^2\chi_k)dA. \tag{13}$$

C. For panel *b*

Assume that the ith external exciting point force is applied at position σ_i on panel b with a direction towards enclosure. Note that the normal displacement w_b is positive outward; we can express the motion equation of panel b as

$$m_b\ddot{w}_b+D_b\nabla^4(w_b)=p(\xi_b,t)-\sum_l F(\sigma,t)\,\delta(\sigma-\sigma_l), \tag{14}$$

where $\delta(\cdot)$ is the two-dimensional Dirac function, $F(\sigma,t)$ is applied force distribution. Via the same processing procedure as before, the coupling motion equation of panel b is obtained as

$$\ddot{q}_s^b+\omega_s^{b^2}q_s^b=\frac{A_b}{M_s^b}\sum_n p_nC_{ns}^b-\frac{1}{M_s^b}\sum_l \sigma_{sl}F_l, \tag{15}$$

where

$$\sigma_{sl}=\varphi_s^b(\sigma_l). \tag{16}$$

The internal dissipation mechanisms in panels and enclosure fluid can be taken into account by including viscous damping terms in the modal Eqs. (8), (12), and (15). Hence, we can get the following equations:

$$\ddot{p}_n+2\zeta_n\omega_n\dot{p}_n+\omega_n^2p_n=\frac{\rho_0c_0^2A_a}{M_n}\sum_r C_{nr}^a\ddot{q}_r^a - \frac{\rho_0c_0^2A_b}{M_n}\sum_s C_{ns}^b\ddot{q}_s^b, \tag{17}$$

$$\ddot{q}_r^a+2\zeta_r^a\omega_r^a\dot{q}_r^a+\omega_r^{a^2}q_r^a=\frac{A_a}{M_r^a}\sum_n p_nC_{nr}^a-\frac{1}{M_r^a}\sum_k \epsilon_{rk}V_k, \tag{18}$$

$$\ddot{q}_s^b+2\zeta_s^b\omega_s^b\dot{q}_s^b+\omega_s^{b^2}q_s^b=\frac{A_b}{M_s^b}\sum_n p_nC_{ns}^b-\frac{1}{M_s^b}\sum_l \theta_{sl}F_l, \tag{19}$$

where ζ_n, ζ_r^a, and ζ_s^b are modal damping ratios of fluid and panels, respectively, which are usually determined based upon empirical data.

Considering the case for a harmonic external disturbance with frequency ω, we can express Eqs. (17), (18), and (19) in the form of matrices and time-independent complex amplitude vectors

$$\{P\}=\lfloor Z^a\rfloor\{Q^a\}+\lfloor Z^b\rfloor\{Q^b\}, \tag{20}$$

$$\{Q^a\}=\lfloor H^a\rfloor\{P\}-\lfloor H^V\rfloor\{V_{\mathrm{con}}\}, \tag{21}$$

$$\{Q^b\}=\lfloor H^b\rfloor\{P\}-\lfloor H^F\rfloor\{F\}. \tag{22}$$

Combining Eqs. (20), (21), and (22), the time-independent complex amplitude of sound pressure then can be obtained as

$$\{P\}=\{P_F\}+[B]\{V_{\mathrm{con}}\}, \tag{23}$$

where $\{P_F\}$ is the modal complex amplitude of sound pressure in the uncontrolled case, and

$$\{P_F\}=[G]^{-1}[Z^b][H^F]\{F\}, \tag{24}$$

$$[B]=[G]^{-1}[Z^a][H^V], \tag{25}$$

where

$$[G]=(\lfloor Z^a\rfloor\lfloor H^a\rfloor+\lfloor Z^b\rfloor\lfloor H^b\rfloor-I), \tag{26}$$

in which I is an identity matrix. Then, the acoustic pressure in enclosure can be written as

$$p(\omega)=\{\Psi\}^T\{P_F\}+\{\Psi\}^T[B]\{V_{\mathrm{con}}\} =p_F(\omega)+\{\Psi\}^T[B]\{V_{\mathrm{con}}\}, \tag{27}$$

where $p_F(\omega)$ is the sound pressure in the uncontrolled case.

III. CONTROL SCHEMES

In this section, two different cost functions are presented, along with the associated optimal control inputs and the system responses. Each of the schemes presented below can be used as a basis for a feedforward control scheme.

A. Scheme 1

The temporal and spatial average of the sound pressure squared in enclosure is given by

$$\langle\overline{p^2}\rangle_{\mathrm{gl}}=\frac{1}{2V}\int_V pp^*\,dV, \tag{28}$$

where the subscript gl denotes global control, and asterisk denotes complex conjunction. From Eq. (27), we can obtain

$$\langle\overline{p^2}\rangle_{\mathrm{gl}}=\{V_{\mathrm{con}}\}^H[a_1]\{V_{\mathrm{con}}\}+\{V_{\mathrm{con}}\}^H[b_1] +[b_1]^H\{V_{\mathrm{con}}\}+[c_1], \tag{29}$$

where

$$[a_1]=[B]^H[\Delta][B], \tag{30}$$

$$[b_1]=[B]^H[\Delta]\{P_F\}, \tag{31}$$

$$[c_1]=\{P_F\}^H[\Delta]\{P_F\}, \tag{32}$$

in which the superscript H indicates complex conjugate transpose, $[\Delta]$ is a diagonal matrix with the ith element being $(1/2V)\int_V\psi_i^2dV$. Equation (29) shows that the cost function is a quadratic function of control inputs. Hence, the optimal values for control voltage amplitudes and mean-square pressure are given by

$$\{V_{\mathrm{con}}\}=-[a_1]^{-1}[b_1], \tag{33}$$

$$\langle\overline{p^2}\rangle_{\mathrm{gl}}^{\mathrm{opt}}=[c_1]-[b_1]^H[a_1]^{-1}[b_1]. \tag{34}$$

B. Scheme 2

In order to implement a global sound-pressure control, an infinite number of error microphones is needed inside the enclosure. This is obviously impractical. Hence, for intended practicality, an alternative is to use N number of sensors to perform a local minimization. Thus, we define the local optimal objective function as

$$\langle\overline{p^2}\rangle_{\mathrm{loc}}=\frac{1}{2N}\{P\}^H[\Psi_N][\Psi_N]^T\{P\}, \tag{35}$$

where T denotes transpose, $[\Psi_N]$ is a matrix of mode shapes at the N sensor locations, and the subscript loc denotes local control. We can see that when N tends to infinity, the control will tend to global control. The optimal values for control voltage amplitudes and mean-square pressure are found to be

$$\{V_{\mathrm{con}}\}=-[a_2]^{-1}[b_2], \tag{36}$$

$$\langle\overline{p^2}\rangle_{\mathrm{loc}}^{\mathrm{opt}}=[c_2]-[b_2]^H[a_2]^{-1}[b_2], \tag{37}$$

where

$$[a_2]=(1/2N)[B]^H[\Psi_N][\Psi_N]^T[B], \tag{38}$$

$$[b_2]=(1/2N)[B]^H[\Psi_N][\Psi_N]^T\{P_F\}, \tag{39}$$

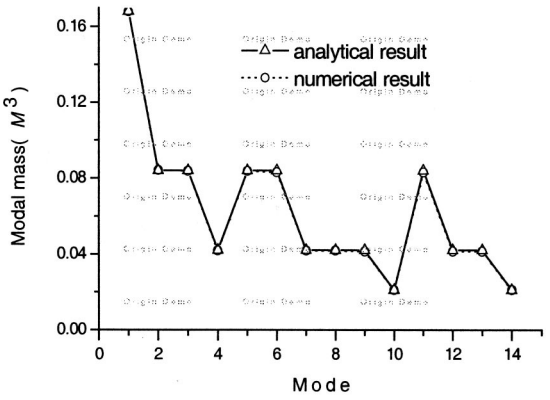


FIG. 2. Comparisons of modal masses of cuboid.

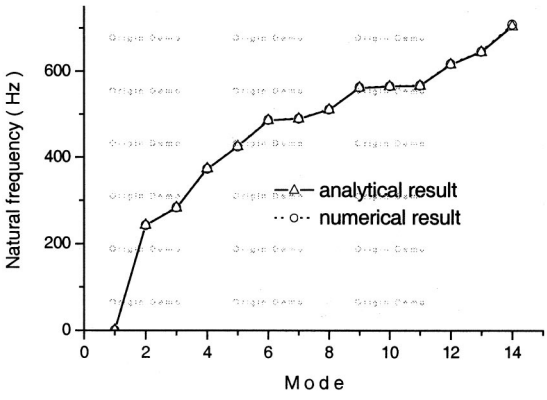


FIG. 3. Comparisons of natural frequencies of cuboid.

$$[c_2]=(1/2N)\{P_F\}^H[\Psi_N][\Psi_N]^T\{P_F\}. \tag{40}$$

IV. CALCULATIONS OF MODAL COUPLING COEFFICIENTS

The calculations of modal coupling coefficients are crucial for an irregular three-dimensional enclosure. As for regular enclosures, we could theoretically determine the coupling coefficients and modal masses according to the mode expressions of enclosures. But, for the enclosure shown in Fig. 1, the acoustic modal shapes can only be determined by numerical calculations. The numerical results of the modal shapes can then be used to calculate the modal coupling coefficients and modal masses, which are used as substitutes of theoretical values in the above motion equations. In order to verify the numerical calculations, a regular cavity with dimensions of $600\times400\times700$ mm is considered. The nth modal frequencies, mode shapes, and modal masses of this regular cuboid can be expressed, respectively, as

$$f_n=\frac{c_0}{2}\sqrt{\left(\frac{n_x}{L_x}\right)^2+\left(\frac{n_y}{L_y}\right)^2+\left(\frac{n_z}{L_z}\right)^2}, \tag{41}$$

$$\psi_n=\cos\frac{n_x\pi}{L_x}\cos\frac{n_y\pi}{L_y}\cos\frac{n_z\pi}{L_z}, \tag{42}$$

$$M_n=\int_V\psi_n^2\,dV. \tag{43}$$

According to the above three expressions we can get the modal parameters by analytical calculations, whereas the corresponding numerical data of modal parameters are possible as well. Figures 2 and 3 illustrate the comparisons of modal masses and modal frequencies obtained by the two methods, respectively. It can be found from the two figures that the numerical results are the same as that obtained by analytical equations, i.e., the numerical method used here to determine the modal parameters is credible.

Using the same numerical calculation procedures used in the above section, we now consider the irregular enclosure shown in Fig. 1 with the dimensions of $L_{xa}=600$ mm, $L_{xb}=750$ mm, $L_y=450$ mm, and $L_z=750$ mm. The first 17 modal frequencies and modal masses are calculated. The two panels' materials are aluminum with the same thickness of 1 mm. The first 52 modes for panel a and first 61 modes for panel b are also calculated, respectively, according to theoretical mode expressions. Then, the modal coupling coefficients can be calculated based on these results. Table I illustrates the first several modal frequencies of enclosure and two panels. They may be referred in the following mechanisms analyses.

V. RESULTS AND DISCUSSIONS

For numerical simulation predictions, the sound-pressure level in enclosure is characterized by

$$\mathrm{SPL}=20\log_{10}\left(\frac{\sqrt{\langle\overline{p^2}\rangle}}{20\times10^{-6}}\right), \tag{44}$$

where the reference pressure is taken as 20 μ Pa. In Fig. 1, two panels are made from 1-mm-thickness aluminum plate with the material properties of Young's modulus $E=7.1\times10^{10}$ N/m^2, mass density $\rho=2.7\times10^3$ kg/m^3, and Poisson's ratio $\nu=0.3$. The PZT patches, each having the dimensions of 50×38 mm and thickness of 0.25 mm, with material properties of Young's modulus $E_p=7.24\times10^{10}$ N/m^2, mass den-

TABLE I. The modal indices and modal frequencies of two panels and the first five acoustic natural frequencies of enclosure.

Panel a		Panel b		Enclosure	
Mode	f (Hz)	Mode	f (Hz)	Mode	f (Hz)
1,1	18.81	1,1	16.37	1	0.00
2,1	39.12	2,1	29.37	2	225.54
1,2	54.92	3,1	51.04	3	256.19
3,1	72.98	1,2	52.48	4	345.62
2,2	75.23	2,2	65.48	5	382.40
3,2	109.10	4,1	81.37		
1,3	115.11	3,2	87.15		
4,1	120.37	1,3	112.67		
2,3	135.42	4,2	117.48		
4,2	156.49	5,1	120.37		
3,3	169.28	2,3	125.67		
5,1	181.31	3,3	147.34		
1,4	199.37	5,2	156.49		
4,3	216.67	6,1	168.04		
5,2	217.42	4,3	177.67		
2,4	219.68	1,4	196.93		
3,4	253.54	6,2	204.15		
6,1	255.79	2,4	209.93		
5,3	277.61	5,3	216.67		
6,2	291.91	7,1	224.37		

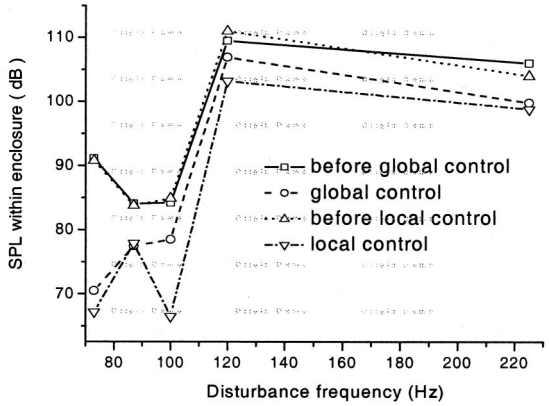


FIG. 4. SPL within enclosure at five selected frequencies.

sity $\rho_p = 2.68 \times 10^3$ kg/m^3, Poisson's ratio $\nu_p = 0.3$, and dielectric constant $d_{31} = 2.7 \times 10^{-10}$ m/V are bound symmetrically to the top and bottom of panel *a*. In simulation and mechanism analyses, one pair of PZT patches, one point force, and two sensors within enclosures are considered. The locations of point force and PZT patches are chosen close to the centers of two panels, respectively. One sensor is at the central part of the enclosure and the other is close to one corner of the enclosure. The center of PZT pair is located on panel *a* at (300,220) mm. The point force on panel *b* is located at $\sigma_1 = (370,220)$ mm, and the two sensors are located at $s_1 = (360,270,375)$ mm and $s_2 = (520.83,375,625)$ mm within the enclosure, respectively. The mass density of air and speed of sound are $\rho_0 = 1.21$ kg/m^3 and $c_0 = 344$ m/s, respectively.

The modal damping ratios of fluid in enclosure and panels are taken to be $\zeta_n = 0.005$ and $\zeta = 0.002$, respectively. According to modal coupling theory, the number of enclosure modes and panel modes should be taken in such a way that the highest modal frequencies of both enclosure modes and panel modes within the frequency range of interest are equal or close, and within one-third of frequency range the results are more acceptable. Hence, they are taken to be 17, 52, and 61, respectively, and the corresponding highest modal frequencies are 695.75, 689.14, and 698.17 Hz, respectively. In the following analyses, the amplitude of point force is $0.3N$, and the disturbance frequencies are chosen as 73, 87, 100, 120, and 225 Hz, respectively. The predictions of SPL within enclosure under these frequencies are illustrated in Fig. 4. Referring to the definitions of two cost schemes, we know that the global control scheme is the temporal and spatial average of the sound pressures for all the points within the enclosure, and the local control scheme is the average of N points within the enclosure. They are different cost functions; therefore, the reductions of SPL in enclosure are different as shown in Fig. 4.

Figure 5 shows the amplitudes and the relative phases of the modal coordinates of panel *a*, panel *b*, and the enclosure when the disturbance frequency is chosen as 73 Hz for the two cost functions. The disturbance frequency corresponds to the (3,1) natural frequency of panel *a*. It can be seen from the Fig. 5 that when the panel *b* is disturbed at 73 Hz and there is no control voltage applied to the PZT patches on panel *a*, the fourth mode of panel *a* is excited, and the amplitude of this dominant mode is much greater than any other modes of two panels before control. From Eqs. (20), (21), and (22), we can obtain

$$\{Q^a\} = \{Q_F^a\} + \{R^a\}\tilde{V}_{\text{con}}, \tag{45}$$

$$\{Q^b\} = \{Q_F^b\} + \{R^b\}\tilde{V}_{\text{con}}, \tag{46}$$

where $\{Q_F^a\}$ and $\{Q_F^b\}$ are the values without control, and

$$\{Q_F^a\} = \lfloor H^a \rfloor \{P_F\}, \tag{47}$$

$$\{Q_F^b\} = \lfloor H^b \rfloor \{P_F\}, \tag{48}$$

$$\{R^a\} = \lfloor H^a \rfloor \{B\} - \{H^V\}, \tag{49}$$

$$\{R^b\} = \lfloor H^b \rfloor \{B\} - \{H^F\}. \tag{50}$$

It can be seen from Eqs. (47), (48) that when the disturbance frequency is near the fourth natural frequency of panel *a*, the values of the elements in the fourth row in matrix $\lfloor H^a \rfloor$ must be much greater than any others in matrixes $\lfloor H^a \rfloor$ and $\lfloor H^b \rfloor$; as a result, the fourth mode of the panel *a* becomes the dominant mode. When control voltages are applied, the amplitude of the fourth mode of panel *a* is suppressed under each of the two control schemes. At the same time, several mode amplitudes of panel *b* become much greater than before control and their relative phases vary. Although many modal amplitudes of panel *b* increase after control, most acoustic modes of enclosure are suppressed, especially the first one. Recalling Eq. (20), we can obtain the *j*th modal coordinate of enclosure

$$\tilde{p}_j = \sum Z_{ji}^a \tilde{q}_i^a + \sum Z_{ji}^b \tilde{q}_i^b. \tag{51}$$

It can be seen from above equation and the definitions of $\{Q^a\}$ and $\{Q^b\}$ that when the control voltages are applied, the variations of the amplitudes and phases of $\tilde{q}_i^a$ and $\tilde{q}_i^b$ result in the variations of acoustic mode coordinates. It can be seen from Fig. 4 that at this frequency the SPL within enclosure is reduced by 20.6 and 22.4 dB with global and local control, respectively. Obviously the control mechanism for panel *a* is modal amplitude suppression and for panel *b* is modal restructuring at this frequency.

Figure 6 shows the modal coordinate responses when the disturbance frequency is chosen as 87 Hz, which corresponds to the (3,2) natural frequency of panel *b*. Before control, the seventh mode of panel *b* is the dominant mode, though it is not remarkable in the figure. After the control, the amplitude of this mode becomes much greater, noticing the coordinate scale, and the phase is changed. Meanwhile, the amplitudes of several modes of panel *a* become much greater than before the control, especially the fourth and seventh modes. It can be seen from Fig. 6 that their phase also changes greatly, as do several modes of enclosure. In this situation, the variations of amplitudes and phases of two panels cooperate in the system, and more acoustic modes would contribute sound pressure within the enclosure with the result of noise attenuation. At this frequency, it can be concluded that the modal restructuring plays a main role in both panels.

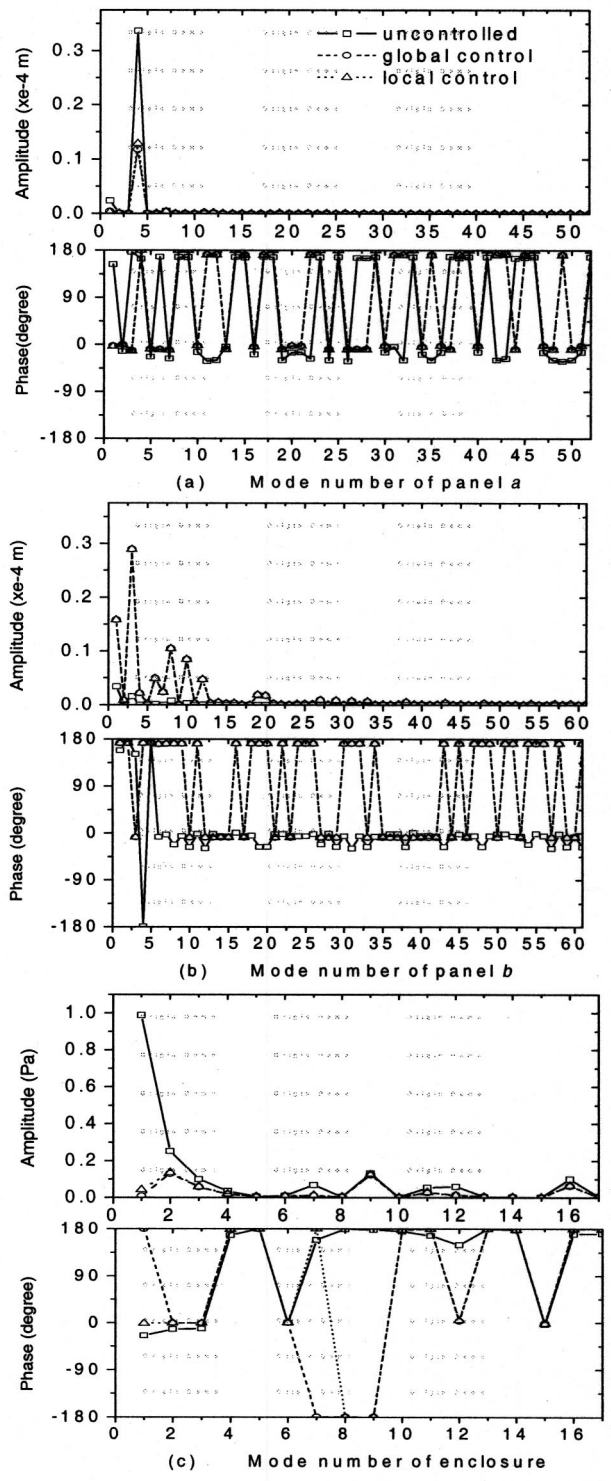


FIG. 5. Modal coordinate responses at 73 Hz (a) amplitudes and phases of panel *a*; (b) amplitudes and phases of panel *b*; (c) amplitudes and phases of enclosure.

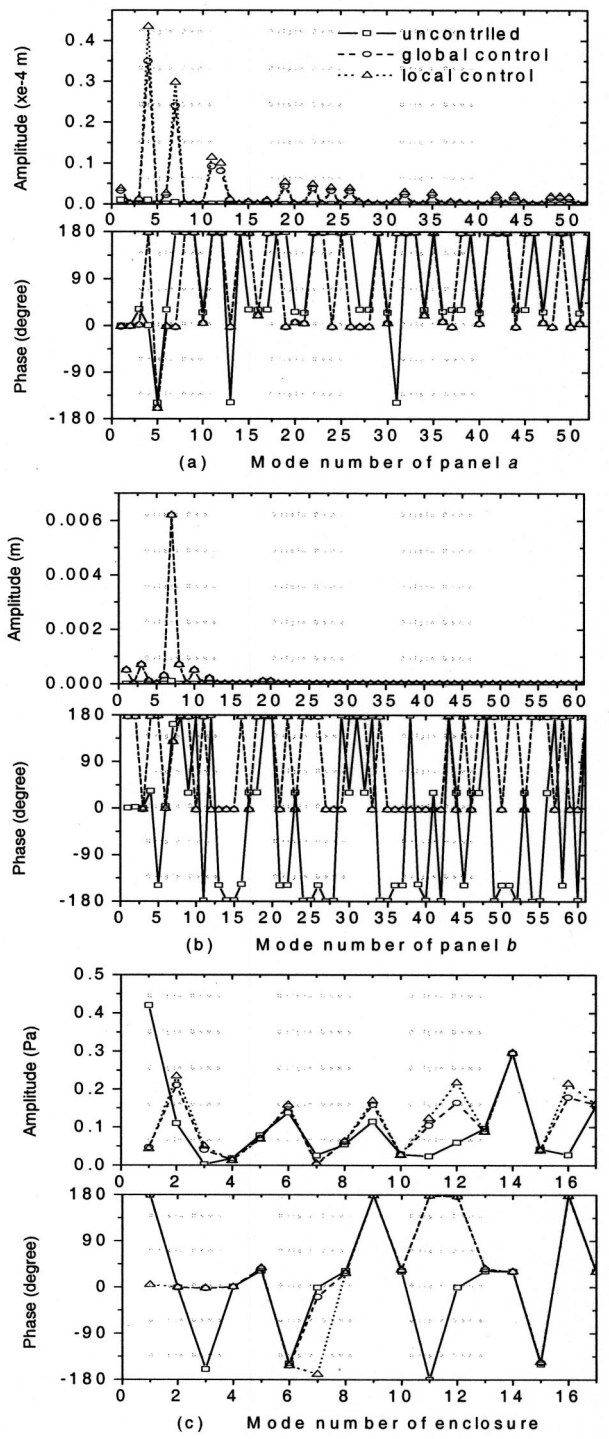


FIG. 6. Modal coordinate responses at 87 Hz (a) amplitudes and phases of panel *a*; (b) amplitudes and phases of panel *b*; (c) amplitudes and phases of enclosure.

The mode coordinate responses are shown in Fig. 7 when the disturbance frequency is 100 Hz, which is a frequency that does not correspond to any of the natural frequencies of two panels and the enclosure. Several modes are increased after control both for panel *a* and for panel *b*, and their phases likewise are changed. After the control voltages are applied, the first acoustic mode is suppressed and the second increased. The SPL within enclosure is reduced as shown in Fig. 4. It can be elucidated that the noise attenuation is caused by modal restructuring for two panels at 100 Hz.

However, only one dominant panel mode is increased in panel *b* at 120 Hz [corresponding to (4,1) natural frequency of panel *a* and (5,1) natural frequency of panel *b* at the same time] after control voltages are applied, as shown in Fig. 8, which is the resonant mode of the panel *b*. The amplitude of this mode is remarkable. It can be seen from the figure that

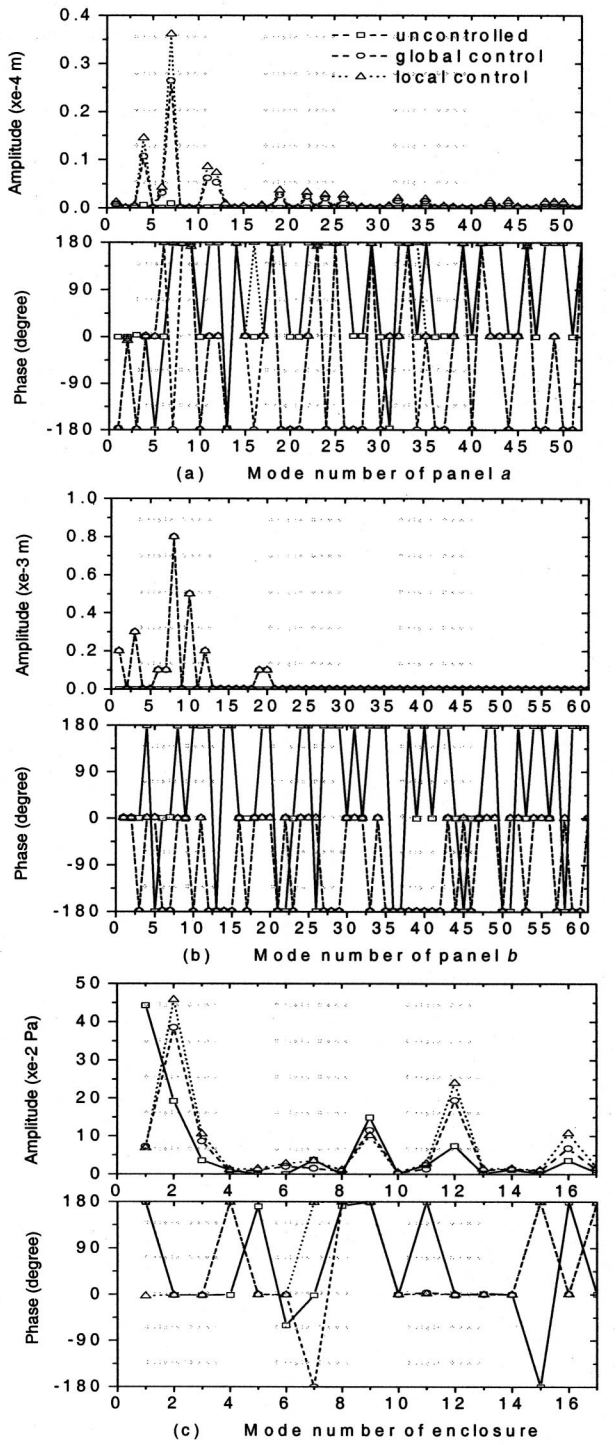


FIG. 7. Modal coordinate responses at 100 Hz (a) amplitudes and phases of panel *a*; (b) amplitudes and phases of panel *b*; (c) amplitudes and phases of enclosure.

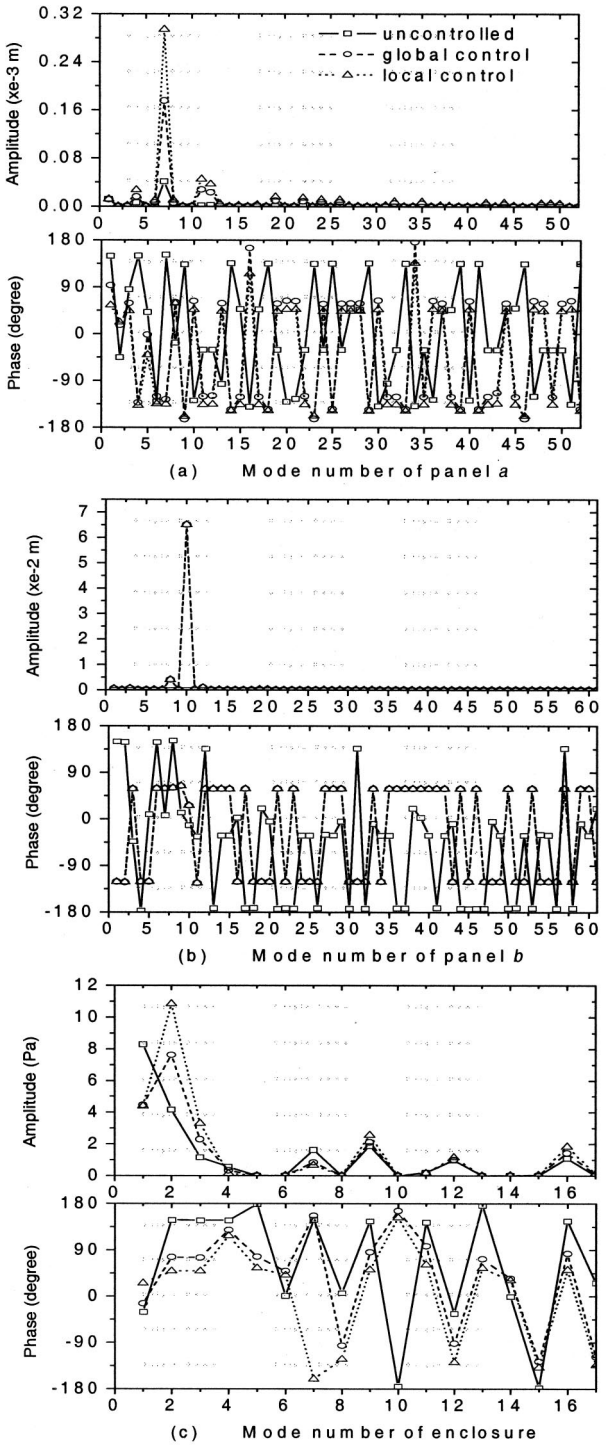


FIG. 8. Modal coordinate responses at 120 Hz (a) amplitudes and phases of panel *a*; (b) amplitudes and phases of panel *b*; (c) amplitudes and phases of enclosure.

the phase and amplitude of dominant mode of panel *a* are changed as well. In this situation, the appliance of control voltages excites the movements of two panels further. But, the SPL within the enclosure is also decreased by 2.9 and 7.3 dB with the global control and local control scheme, respectively. Therefore, the SPL within the enclosure is reduced by modal restructuring for two panels at this disturbance frequency.

When the disturbance frequency is chosen at 225 Hz, which corresponds to the second acoustic natural frequency of enclosure, the system responses are shown in Fig. 9. After controlling, many modal amplitudes of panel *a* are increased, only one panel mode in panel *b* is increased, and the phases of the dominant modes of two panels are changed at the same time. Within the enclosure, the second acoustic mode is excited before control and is suppressed after control. Therefore, the modal restructuring also works at this disturbance frequency for two panels.

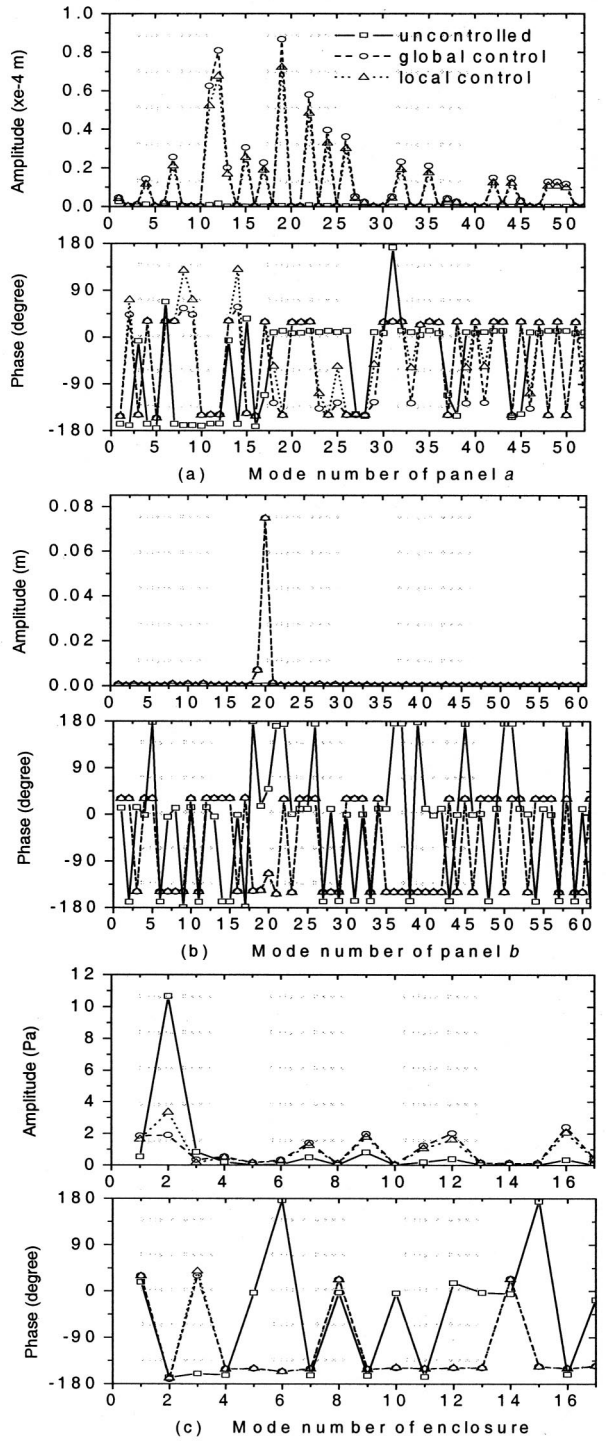


FIG. 9. Modal coordinate responses at 225 Hz (a) amplitudes and phases of panel *a*; (b) amplitudes and phases of panel *b*; (c) amplitudes and phases of enclosure.

VI. CONCLUSIONS

This paper has developed a new modeling method for active minimization of noise within an irregular enclosure, which is proven credible and not subject to the limitation of enclosure boundaries. Numerical investigations demonstrate that good results are obtained in minimizing SPL within irregular enclosure with this method and two cost functions.

It is found that there are different control mechanisms at different disturbance frequencies. Two control mechanisms are demonstrated in this paper. When the disturbance frequency is chosen close to one of the natural frequencies of panel *a*, the amplitude of the dominant mode of panel *a* is suppressed, and the modes of panel *b* are restructured at the same time. When the disturbance frequency is chosen close to one of the natural frequencies of panel *b*, the modes of two panels are all restructured and the SPL in enclosure is reduced.

[1] L. J. Eriksson, "Recent trend in the development of active sound and vibration control systems," Proceedings of Noise-Con 94, Ft. Lauderdale, Florida, 271–278 (1994).

[2] C. R. Fuller, "Recent advances in active noise control," AIAA J. **29**, 1058–1066 (1991).

[3] P. A. Nelson, A. R. D. Curtis, and S. J. Bullmore, "The active minimization of harmonic enclosed sound field. I. Theory," J. Sound Vib. **117**, 1–13 (1987).

[4] A. J. Bullmore, A. R. D. Curtis, and S. J. Elliot, "The active minimization of harmonic enclosed sound field. II. A computer simulation," J. Sound Vib. **117**, 15–33 (1987).

[5] S. J. Elliot, A. R. D. Curtis, A. J. Bullmore, and P. A. Nelson, "The active minimization of harmonic enclosed sound field. III. Experimental verification," J. Sound Vib. **117**, 35–38 (1987).

[6] X. J. Qiu, "Mechanisms of active control of noise transmission through a panel into a cavity using point force actuator on the panel," J. Sound Vib. **182**, 167–170 (1995).

[7] K. Seto, M. Z. Ren, and F. Doi, "Feedback vibration control of a flexible plate at audio frequencies by using a physical state-space approach," J. Acoust. Soc. Am. **103**, 924–934 (1998).

[8] A. Sampath and B. Balachandran, "Studies on performance functions for interior noise control," Smart Mater. Struct. **6**, 315–332 (1997).

[9] H. T. Banks, "Modeling and control of acoustic structure interaction problems via piezoceramic actuators: 2-D numerical example," ASME J. Vibr. Acoust. **116**, 386–396 (1994).

[10] S. Koshigoe, A. Teagle, and A. Gordon, "A time domain study of active control of sound transmission due to acoustic pulse excitation," J. Acoust. Soc. Am. **97**, 313–323 (1995).

[11] S. Koshigoe and J. W. Murdock, "A new approach for active control of sound transmission through an elastic plate backed by a rectangular cavity," J. Acoust. Soc. Am. **94**, 900–907 (1993).

[12] X. Q. Bao, V. V. Varadan, and V. K. Varadan, "Active control of sound transmission through a plate using a piezoelectric actuator and sensor," Smart Mater. Struct. **4**, 231–239 (1995).

[13] S. Poh, A. Baz, and B. Balachandran, "Experimental adaptive control of sound radiation from a panel into an acoustic cavity using active constrained layer damping," Smart Mater. Struct. **5**, 649–659 (1996).

[14] W. Shields, J. Ro, and A. Baz, "Control of sound radiation from a plate into an acoustic cavity using active piezoelectric-damping composites," Smart Mater. Struct. **7**, 1–11 (1998).

[15] S. Kawamura, "Analysis in required frequency band of electric-structural-acoustic coupled system," JSME Int., J., Ser. C **64**, 6–13 (1998).

[16] K. S. Sum and J. Pan, "An analytical model for bandlimited response of acoustic–structural coupled systems. I. Direct sound field excitation," J. Acoust. Soc. Am. **103**, 911–923 (1998).

[17] F. Fahy, *Sound and Structural Vibration: Radiation, Transmission, and Response* (Academic, London, 1985), Chap. 6, pp. 241–268.

[18] J. Pan, C. H. Hansen, and D. A. Bies, "Active control of noise transmission through a panel into a cavity. I. Analytical study," J. Acoust. Soc. Am. **87**, 2098–2108 (1990).

[19] J. Pan and C. H. Hansen, "Active control of noise transmission through a panel into a cavity. II. Experimental study," J. Acoust. Soc. Am. **90**, 1488–1492 (1991).

[20] X. J. Qiu, J. Z. Sha, and J. Yang, "Mechanisms of active control of noise transmission through a panel into a cavity using a point force actuator on the panel," J. Sound Vib. **182**, 167–170 (1995).

[21] S. Koshigoe and J. W. Murdock, "A united analysis of both active and passive damping for a plate with piezoelectric transducers," J. Acoust. Soc. Am. **93**, 346–355 (1993).

A hybrid SEA/modal technique for modeling structural-acoustic interior noise in rotorcraft

V. Jayachandran and M. W. Bonilha[a)]
United Technologies Research Center, East Hartford, Connecticut 06108

(Received 3 January 2002; revised 27 November 2002; accepted 20 December 2002)

This paper describes a hybrid technique that combines Statistical Energy Analysis (SEA) predictions for structural vibration with acoustic modal summation techniques to predict interior noise levels in rotorcraft. The method was applied for predicting the sound field inside a mock-up of the interior panel system of the Sikorsky S-92 helicopter. The vibration amplitudes of the frame and panel systems were predicted using a detailed SEA model and these were used as inputs to the model of the interior acoustic space. The spatial distribution of the vibration field on individual panels, and their coupling to the acoustic space were modeled using stochastic techniques. Leakage and nonresonant transmission components were accounted for using space-averaged values obtained from a SEA model of the complete structural-acoustic system. Since the cabin geometry was quite simple, the modeling of the interior acoustic space was performed using a standard modal summation technique. Sound pressure levels predicted by this approach at specific microphone locations were compared with measured data. Agreement within 3 dB in one-third octave bands above 40 Hz was observed. A large discrepancy in the one-third octave band in which the first acoustic mode is resonant (31.5 Hz) was observed. Reasons for such a discrepancy are discussed in the paper. The developed technique provides a method for modeling helicopter cabin interior noise in the frequency mid-range where neither FEA nor SEA is individually effective or accurate. [DOI: 10.1121/1.1548157]

PACS numbers: 43.40.At, 43.20.Tb, 43.40.Rj [JHG]

I. INTRODUCTION

The prediction and control of interior noise levels in helicopters is an important problem. The principal sources of noise are the rotor (low frequency aerodynamically generated noise), gearbox (strongly tonal gear clash noise), turbulent boundary layer excitation (broadband noise due to airflow past the helicopter in forward flight), and miscellaneous sources (engine, hydraulic systems, etc.).[1] The modeling of these sources is a complex undertaking and much work has been performed in this area.

Once the sources have been quantified, the transmission path must be modeled accurately in order to predict the radiated sound levels inside the cabin. In a typical helicopter, the vibrating trim panels that enclose the cabin generate a large component of the noise field. The transmission path to these panels is via the airframe structure and the trim cavities of the helicopter. The modeling of this transmission path can be performed using various techniques, depending on the frequency range of application. In the low frequency domain, popular methods for modeling cabin interior noise include Finite Element Analysis (FEA), and modal expansion type of techniques. Modal expansion techniques work well for simple geometries, are easy to implement, and are numerically efficient. Their use of eigenfunctions helps retain the physics of the problem in the solution process, facilitating a better understanding of the system behavior. FEA relies on using simpler basis functions for determining the response of systems with complex geometry. However, both techniques can get computationally extremely expensive and also somewhat inaccurate with increasing frequency as the mode count goes up.

At higher frequencies, Statistical Energy Analysis (SEA) is a powerful numerical solution procedure. SEA tracks the energy balance in a system and provides frequency and space averaged measures of the response.[2] A major advantage of SEA results from grouping resonant modes together and treating their response statistically. This allows for a significant reduction in the number of degrees of freedom in the model. The disadvantage is that it is not possible to predict spatial variations in the response. By its very nature, the technique relies on averaging in the space and frequency domains to derive a solution. SEA has been used in the past for modeling the high frequency structural dynamics and interior noise of rotorcraft.[3–6]

FEA or SEA may be chosen for modeling a system depending on whether the response of the components is modal or diffuse. In a helicopter, however, this choice is not very clear. Even at low to mid frequencies (50–400 Hz), the interior panels can have high modal densities. SEA models have been shown to perform very well in predicting the structural response of these helicopter subsystems to very low frequencies. On the other hand, the mode count of the large interior acoustic space can be quite low in this frequency range. The high modal densities of the panels, coupled with the low modal density of the interior acoustic space, makes it difficult to use a purely FEA or SEA based technique for predicting the complete structural-acoustic response of the system. Of course, once the excitation fre-

[a)]Author to whom correspondence should be addressed; electronic mail: bonilhmw@utrc.utc.com, Tel: 860-610-7528.

quency is well above the interior acoustic space's Schroeder frequency, the whole system can be modeled using SEA. It has been shown that an exclusively SEA-based model breaks down below 400 Hz in modeling *acoustic noise* inside *structurally* excited helicopter cabin systems.[7] This will also be shown to be true (in a later section) for the system being modeled in this paper. Hence it is important to investigate a mid-frequency model that can bridge the gap between predictions that may be accurately obtained from purely FEA and SEA based models.

If the results from an SEA-based structural response model can be somehow coupled with a modal technique for predicting the interior acoustic response, then the resultant hybrid model would be able to bridge the aforementioned gap in analysis techniques. Bonilha and Fahy developed such a hybrid technique and demonstrated it on a very simple system.[8] This paper presents an extension of their methodology to more complex situations along with an actual industrial application.

Hybrid models have been investigated recently by a number of other researchers. Specifically, in the structural-acoustic modeling domain, Shorter[9] and Langley and Bremner[10] have been working to develop general hybrid modeling software in which subsystems can be classified as deterministic or fuzzy in different frequency bands depending on their mode counts. The key problem in their analysis is similar to ours—that of classifying the subsystems and specifying the coupling at their interfaces. References 11–14 describe some other work that combine SEA and FEA techniques for different types of problems, including rotorcraft and aircraft interior noise.

The current work extends the methodology in the article by Bonilha and Fahy[8] to the rotorcraft interior noise problem. The problem considered in Ref. 8 was that of a rectangular enclosure excited by one vibrating wall. The current paper further develops the methodology to consider partially vibrating boundaries (baffled sources, which may be off center), multiple boundary sources, and possible relative phasing between sources. Vibrating panels with cut out sections (e.g., trim panels in helicopters with cut-outs for windows) have been treated in a simple manner by using rectangular panels with equivalent radiating areas. The results presented in this paper coupled SEA models for the vibrating interior panels with a modal model of the interior acoustic space. The space and frequency averaged panel vibration amplitudes were used as primary inputs to the interior acoustic system model. Nonresonant transmission and leakage contributions were accounted for by using simple frequency and space averaged corrections obtained from a SEA model of the complete system. The model was validated by comparing its predictions with those measured in a laboratory test under structural excitation. All numerical programming and computations were performed in MATLAB. The following sections of the paper detail the mathematical development of the problem and the results obtained from the study.

II. MATHEMATICAL DEVELOPMENT

The system studied in this paper was the interior panel system and the enclosed acoustic space of the S-92 Helibus

FIG. 1. Sikorsky S-92A Helibus.

in its North Sea Offshore (NSO) configuration (Fig. 1). It measures 18 ft 7 in. (5.67 m) in length, 6 ft 7 in. (2.0 m) in width and 6 ft (1.83 m) in height. The S-92 also has a luggage compartment with 110 cu ft. (3.1 m^3) volume behind the interior aft bulkhead.

The aircraft interior was modeled as a rectangular enclosure with dimensions l_x, l_y, and l_z in the fore/aft, port/starboard, and vertical directions, respectively. The speed of sound in the acoustic medium and its density are mathematically denoted by c_0 and ρ_0, respectively, in the following analysis. The interior acoustic eigenfunctions were obtained assuming hard wall boundary conditions. Let the natural frequencies and modal damping factors be represented by ω_n and η_n, respectively. Then the pressure field at any location $\mathbf{r}$ may be obtained through a modal summation of the form

$$p(\mathbf{r},t)=\rho_0 c_0^2 \sum_n p_n(t)\psi_n(\mathbf{r}), \tag{1}$$

where the dynamics of the nth mode are given by

$$\ddot{p}_n+\eta_n\omega_n\dot{p}_n+\omega_n^2 p_n=\frac{F_n(t)}{\Lambda_n}. \tag{2}$$

Here, $F_n(t)$ is a generalized source term, $\psi_n(\mathbf{r})$ is the nth acoustic eigenfunction, and Λ_n is a modal normalization coefficient. Hence, the complex pressure can be written as

$$p(\mathbf{r},\omega)=\frac{\rho_0 c_0^2}{V}\sum_{j=1}^{NPanels}\sum_n \frac{\psi_n(\mathbf{r})}{\Lambda_n}\frac{X_n-iY_n}{X_n^2+Y_n^2} \times\beta_j\int_{S_j}A_j(\mathbf{r}_j,\omega)\psi_n(\mathbf{r}_j)dS_j, \tag{3}$$

where ω is the frequency, V is the volume of the enclosure, $A_j(\mathbf{r}_j,\omega)$ and S_j are the surface acceleration field and the surface area of the jth panel, $X_n=(\omega_n^2-\omega^2)$ and $Y_n=\omega_n\omega\eta_n$. Note that β_j conveys the relative phase information between the different panels, i.e., this parameter may be used to allow the panels to vibrate with relative phase. This can be important at very low frequencies where opposite panels vibrating in phase can cause a very different acoustic response as compared to two panels working 180 degrees out of phase with each other.

We can now derive the squared pressure magnitude at any location as

$$S_p(\mathbf{r},\omega)=p^*p=\frac{(\rho_0 c_0^2)^2}{V^2}\sum_n\sum_m\frac{\psi_n(\mathbf{r})}{\Lambda_n}\frac{\psi_m(\mathbf{r})}{\Lambda_m}\times\left[\frac{X_n-iY_n}{X_n^2+Y_n^2}\right]\left[\frac{X_m+iY_m}{X_m^2+Y_m^2}\right]\times\sum_{j=1}^{NPanels}\sum_{i=1}^{NPanels}\beta_j\int_{S_j}A_j^*(\mathbf{r}_j,\omega)\psi_n(\mathbf{r}_j)\times dx_j\,dy_j\beta_i\int_{S_i}A_i(\mathbf{r}_i,\omega)\psi_m(\mathbf{r}_i)dx_i\,dy_i, \tag{4}$$

where p^* is the complex conjugate of p. Also, m and n represent the mth and nth mode number, while i and j are panel indices. We can rewrite the above formula as

$$S_p(\mathbf{r},\omega)=\frac{(\rho_0 c_0^2)^2}{V^2}\sum_{j=1}^{NPanels}\sum_{i=1}^{NPanels}\sum_n\sum_m S_jS_i\frac{\psi_n(\mathbf{r})}{\Lambda_n}\frac{\psi_m(\mathbf{r})}{\Lambda_m}\times\left[\frac{X_n-iY_n}{X_n^2+Y_n^2}\right]\left[\frac{X_m+iY_m}{X_m^2+Y_m^2}\right]C_{nmji}^2(\omega_c), \tag{5}$$

where C_{nmji}^2 represents a structural-acoustic coupling coefficient given by

$$C_{nmji}^2(\omega)=\frac{\beta_jS_{aj}(\omega_c)}{S_j}\frac{\beta_iS_{ai}(\omega_c)}{S_i}\times\int_{S_j}\int_{S_i}\frac{A_j^*(\mathbf{r}_j,\omega)}{S_{aj}(\omega_c)}\frac{A_i(\mathbf{r}_i,\omega)}{S_{ai}(\omega_c)}\times\psi_n(\mathbf{r}_j)\psi_m(\mathbf{r}_i)dS_j\,dS_i. \tag{6}$$

In the above formula, $S_{aj}(\omega_c)$ is the power spectral density of the space and frequency averaged acceleration field on the jth panel, in the band centered at ω_c. The formula for the squared pressure assumes that the coupling coefficients and the space averaged panel vibration velocities vary slowly with frequency; hence if ω lies in a frequency band centered at ω_c, we set $C_{nmji}(\omega)\cong C_{nmji}(\omega_c)$ and $A_j(\mathbf{r}_j,\omega)\cong A_j(\mathbf{r}_j,\omega_c)$.

If we assume that all the panels are in phase, then $\beta_i=\beta_j=1$. If we also assume that the vibration field on different panels is uncorrelated, then $C_{nmji}^2=0$ when i and j are different. This can be justified at high frequencies where the global modes of the aircraft are not expected to propagate and the local modes on the plates dominate the response. This has also been shown to be true at significantly lower frequencies via flight test measurements in rotorcraft. A possible reason is that the panels are not directly connected to each other (all the panels attach separately to the stiff composite frame members via quarter turn fasteners). Hence there is no direct path for the flexural energy to pass from panel to panel. Note that this is an assumption of significant importance—if the vibration field of different panels was assumed to be correlated, then the size of this problem would grow extremely large since cross coupling coefficients would be needed for every pair of panels. If the system contained only two or three panels, this may not have been a big issue. In the current application, we had 43 panels and without the above assumption, this method would have become computationally too expensive. Also note that this assumption is valid for rotorcraft only and may not be valid for other systems enclosed by vibrating panels. Proceeding with the above assumption, we now get a simplified expression for the pressure field given by

$$S_p(\mathbf{r},\omega)=\frac{(\rho_0 c_0^2)^2}{V^2}\sum_{j=1}^{NPanels}\sum_n\sum_m S_j^2\frac{\psi_n(\mathbf{r})}{\Lambda_n}\frac{\psi_m(\mathbf{r})}{\Lambda_m}\times\left[\frac{X_n-iY_n}{X_n^2+Y_n^2}\right]\left[\frac{X_m+iY_m}{X_m^2+Y_m^2}\right]C_{nmj}^2(\omega_c), \tag{7}$$

where the new structural-acoustic coefficient for the jth panel, C_{nmj}^2, is given by

$$C_{nmj}^2(\omega_c)=\frac{S_{aj}^2}{S_j^2}\int_{S_j}\int_{S_j'}\gamma_{aj}(x_j,y_j,x_j',y_j',\omega_c)\times\psi_n(x_j,y_j)\psi_m(x_j',y_j')dS_j\,dS_j'. \tag{8}$$

The function $\gamma_{aj}(x_j,y_j,x_j',y_j',\omega)$ represents the normalized cross power spectral density function on the jth panel.

All the expressions derived to this point have retained the panel vibration fields as completely known variables. However, in the current study, only space and frequency averaged values for the panel accelerations were available (which were computed using an SEA model). The spatial variation of the vibration field on the panels was not available in closed form. Following the approach used by Bonilha and Fahy,[15] the vibration response of any panel can be assumed to consist of a continuous spectrum of structural modes that lie in the wave number space demarcated by the limits of the frequency band under consideration. All the modes are assumed to have the same amplitude, and the wave number space is assumed to be fully populated with modes (this assumption breaks down at very low frequencies where the panel will have very few, discrete modes). We also assumed earlier that each panel vibrates independently from its neighbors and therefore the global modes of the system are not included. Based on these assumptions, we can derive an expression for the normalized cross power spectral density function as

$$\gamma_{aj}(x_j,y_j,x_j',y_j',\omega)=\frac{\int_{\Delta k}\int_0^{\pi/2}X_j(x_j)Y_j(y_j)X_j(x_j')Y_j(y_j')d\theta\,dk_B}{1/S_j\int_{S_j}\int_{\Delta k}\int_0^{\pi/2}X_j^2(x_j)Y_j^2(y_j)d\theta\,dk_B\,dx_j\,dy_j}. \tag{9}$$

In the above expression, k_B and θ represent the bending wave number for the panel in polar coordinates, Δk is the wave number range corresponding to the frequency range under consideration, while $X(x)$ and $Y(y)$ are functions based on Bolotin's dynamic edge effect method for a rectangular plate which are unique for a given boundary condition.[16] Equations (7), (8), and (9) were programmed as

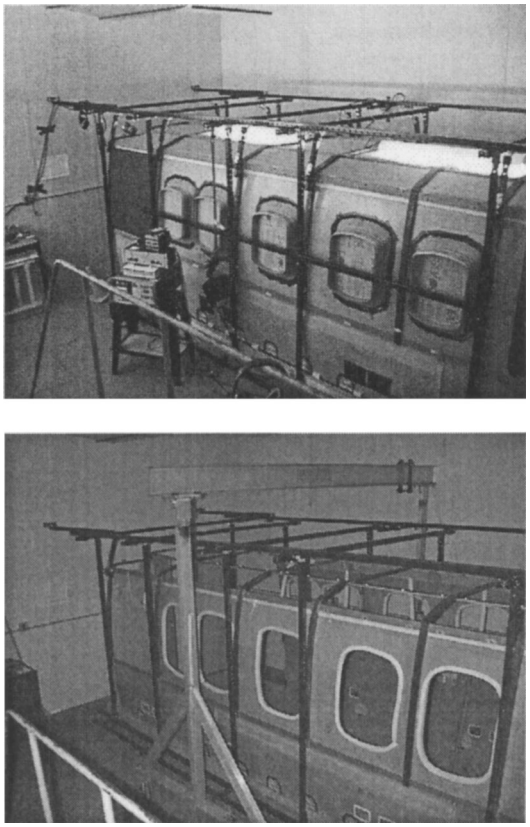


FIG. 2. The S-92 interior panel system mockup at Sikorsky.

a numerical tool in MATLAB. The experiment as well as model are described next.

III. EXPERIMENTAL SETUP

A mock-up of the S-92 interior panel system was assembled inside the Sikorsky reverberation chamber for validating the SEA model of the interior panel system developed at UTRC and Sikorsky.[7] The experimental data gathered were used to derive the SEA parameters of the actual interior so that assumptions in creation of the predictive SEA model could be validated. The development effort for the hybrid model was a result of this validation study where SEA was found to be inaccurate below 400 Hz for predicting structurally excited interior noise. The experimental data from the above validation study was also used for validating the hybrid model.

The S-92 interior uses a modular construction with the panels attached to a hoop frame so that most panels can be easily removed after disengaging a few connectors. This allows easy access to the space between the interior and airframe for repairs or maintenance. In the experimental mockup, the interior panels were attached to the lightweight hoop frame made of composite materials which were in turn connected to a supporting unistrut frame by vibration isolators. The unistrut framing took the place of the airframe in supporting the interior structure. Structural excitation in the form of a shaker attached to the unistrut was applied to the mock-up. Details of the panel system are shown in Figs. 2 and 3. Three sections of the interior mockup were tested in order to verify the dynamic behavior of the various panels,

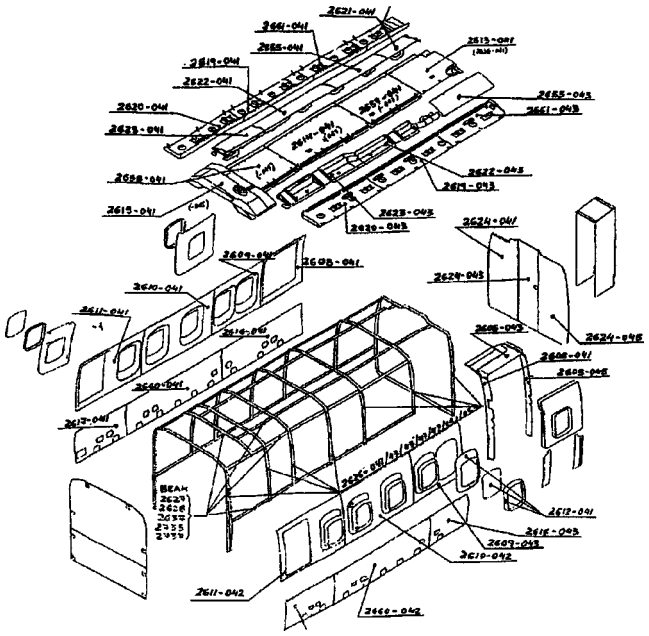


FIG. 3. Different components of the S-92 interior panel system.

beams, and the connections between them. Each section centered on one panel and all connecting beams and neighboring panels were instrumented. Accelerometer data was collected from the panels, whereas, interior noise levels were measured at ten different locations using microphones. A detailed SEA model of the system was developed (in AutoSEA, a commercial software from Vibro-Acoustic Sciences, Inc.), and panel vibration levels were predicted for the different subsystems that were studied. The SEA model was found to perform quite well in predicting the structural response of the system. The model was also used for predicting the interior noise level. The predicted values were observed to be inaccurate below 400 Hz (this is shown in a later figure). Another shortcoming was that the SEA model only provided a space and frequency averaged value for the acoustic pressure and could not be used for studying spatial variations in the sound field at lower frequencies.

IV. NUMERICAL MODEL IMPLEMENTATION

The hybrid model was written using MATLAB. The panel velocity vectors obtained from the SEA model of the interior panel system driven by structural excitation were provided as vibration inputs to the model. Other inputs to the program included the definition of one-third octave bands; an experimentally measured acoustic damping spectrum; material properties, location and size of all the trim panels (made of sandwich composite materials), and windows (made of plexiglass). Nonresonant transmission and leakage contributions to the sound field were added into the hybrid model as simple frequency averaged corrections. These were obtained from the SEA model of the complete structural-acoustic system. The SEA model divided the interior airspace into four sections and hence four sets of correction values were available. However, since these were frequency and space averaged, they did not account for local variations in the sound field within a particular section of the aircraft. To accurately

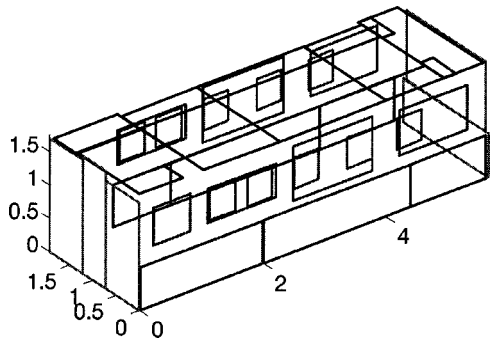


FIG. 4. S-92 interior panel geometry input into the MATLAB program: front view.

model such variations, one would have to study the source and response mechanisms due to acoustic leaks in detail; this was not attempted in the present study.

Figure 4 shows the interior panel geometry of the S-92 aircraft as input into the MATLAB program. The panels were classified into five sets: *upper-side* panels, *lower-side* panels, *end* panels, *ceiling* panels, and *windows*. In all, there were 43 interior panels that were included in the model. Equivalent dimensions for the panels were used in the model, i.e., panels with cutouts were replaced by continuous members centered at the same locations, but with equivalent radiating areas. Some of the panels were slightly curved—for the current study, they were included as flat radiating surfaces. Note that the system boundaries also had a lot of surfaces that did not radiate.

Prior to running the main computation loop to determine the interior acoustic response, the coupling coefficients C_{nmj} between the different panels and acoustic modes were computed using Eqs. (8) and (9) and stored as matrices. This was done because the computation of these coefficients is time-consuming and these values do not change with response location. The first step in the main analysis program was to compute the acoustic modes of the interior space up to a given upper frequency limit and the generation of a modal damping vector for the different modes (using the experimental values). The number and indices of the modes in all the one-third octave bands were determined and stored. These would be used later for summing up the response in each band. The program then ran through a loop over the different one-third octave bands. In each band, the squared pressure at the specified response locations was predicted at 2 Hz intervals using Eq. (7) (for the bands with less than four modes, the response was computed at 1 Hz intervals). This involved computing the modal summation of all the acoustic eigenfunctions at each response location due to all the 43 panels in the system. The squared pressures for all frequency bins within a band were summed to obtain a final one-third octave band value for the acoustic response. The program was used for determining the noise at the ten microphone locations where the data was measured. The results obtained from this procedure are described in the next section.

V. RESULTS

The numerical predictions for the acoustic response were compared with the experimentally measured micro-

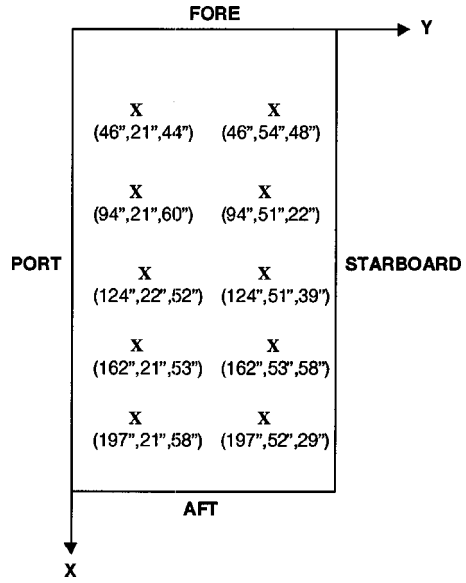


FIG. 5. Plan view of interior microphone locations in the S-92 mockup.

phone data in different sections of the aircraft. Figure 5 shows a plan view of the locations where microphone data was measured in the S-92 mockup (port=left, starboard =right, fore=front, aft=rear). Figures 6–10 show the experimentally measured versus numerically computed one-third octave band sound pressure levels. The first four figures show the mean of the values obtained from the microphones located in individual sections of the aircraft. Figure 10 shows the mean of the values for all the ten microphone positions.

As can be observed, the model grossly overpredicts the response in the 31.5 Hz one-third octave band in the fore and aft locations. This could be because the model assumes hard-walled boundary conditions, whereas the actual S-92 mockup has light walls and large holes in the front and back due to the presence of doors and hatches. Around 30 Hz, the lone dominant longitudinal mode imposes a strong pressure buildup near the ends of the enclosure and hence the model predicts a higher response level. Another possible reason for this inability to correctly predict the response at the first acoustic mode has been reported in the literature. Franchek

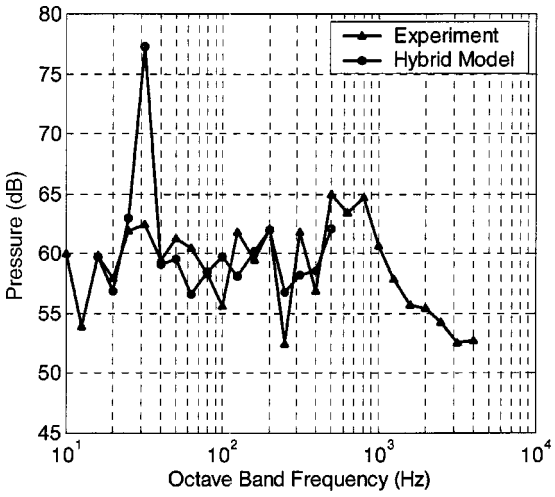


FIG. 6. Mean one-third octave band spectrum for microphones 1 and 2.

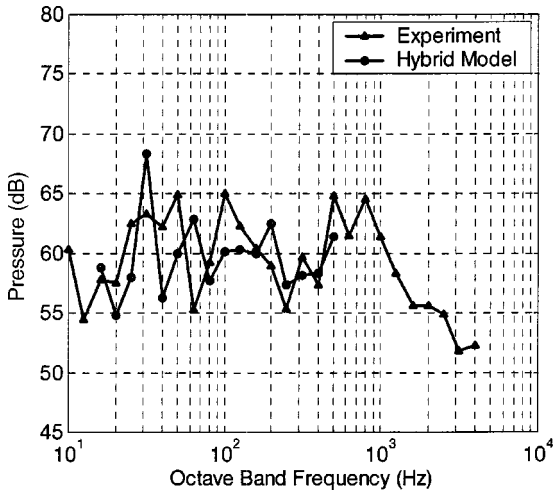


FIG. 7. Mean one-third octave band spectrum for microphones 3 and 4.

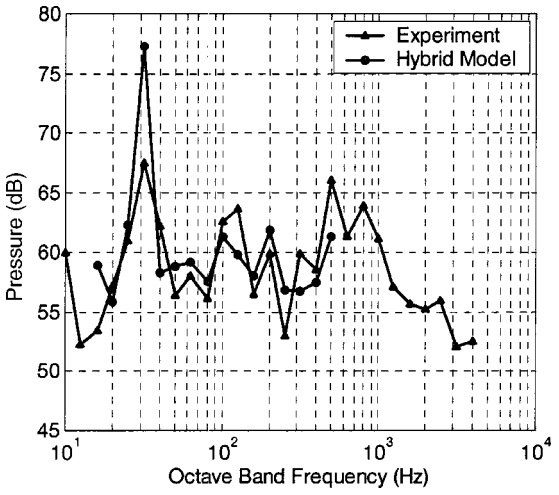


FIG. 9. Mean one-third octave band spectrum for microphones 7, 8, 9, and 10.

and Bernhard[17] employed a FEA model to predict the response inside an acoustic space due to a single vibrating plate using a very fine mesh. A difference of more than 30 dB in the acoustic pressure near the first acoustic resonance between experimental and FEA results was obtained and they observed that slight changes in the model could drastically change the FEA results in the region around the first acoustic mode. This agrees with Pan and Bies[18] conclusion that the overall nature of the coupling in the region of a cavity mode resonance frequency is governed by the average number of plate resonance frequencies available in this region. Therefore, in regions of low plate modal density, small changes in the plate resonance frequencies will lead to substantial changes in the coupling between both systems. Bonilha and Fahy reported a similar phenomenon in their application of the hybrid model.[8] Based on the overall dimensions of all the modeled panels it is expected that only one or two plate modes are interacting with the first longitudinal acoustic mode. Hence the above reasoning could explain the prediction error.

At frequencies above 40 Hz, the predicted pressure levels were typically within 3 dB of the experimentally measured values. A possible source of error in this frequency range is the inaccuracy in the SEA model's prediction of the panel vibration velocities. This arises from the global modes' contributions that are not included in the standard SEA formulation; only local modes of each subsystem are accounted for. Some current research efforts[9,10] are dedicated to the incorporation of global modes in the SEA framework. Errors can also arise if the acoustic modal damping values used in the model are inaccurate. For instance, if the damping is too low, this can lead to the prediction of a stronger response.

An increase in frequency leads to an increase in the modal densities of the acoustic and structural domains and also to an increase in the modal overlap factor. This makes the approach described herein computationally expensive and a purely probabilistic approach is the most appropriate solution. A comparison between the experimental results, hybrid SEA modal, and SEA approach results for the spatial average given by ten microphones is shown in Fig. 10. It is observed that as the frequency increases, the method suggested in this paper starts to approach the SEA results. Above 400 Hz, the SEA model is preferable over the hybrid model

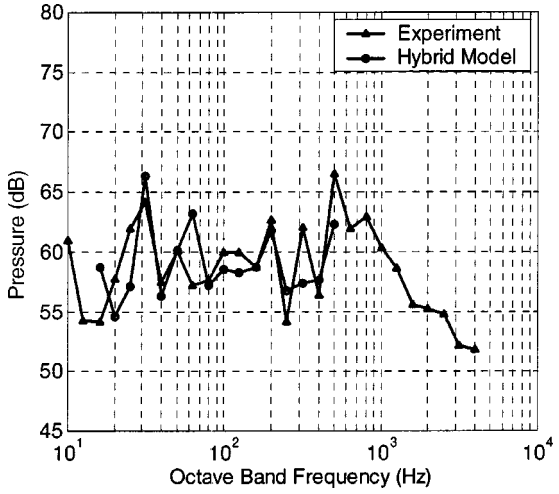


FIG. 8. Mean one-third octave band spectrum for microphones 5 and 6.

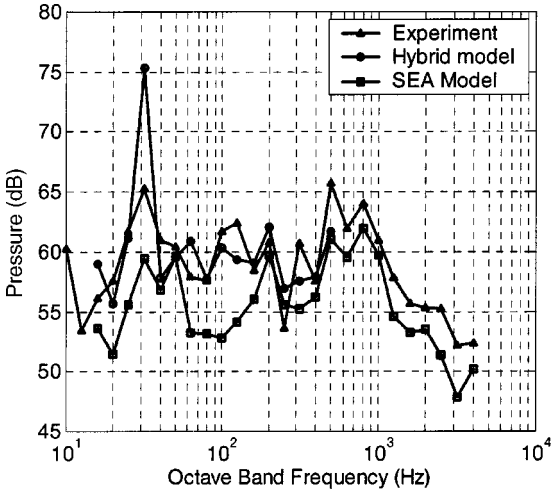


FIG. 10. Mean one-third octave band spectrum for all the microphones in the S-92 mockup. The SEA model results are based on the true space average.

for this application. The SEA model used for this calculation is described in detail in Ref. 7.

The ability of the current approach in predicting the correct spatial response variation was verified by comparing 3D contour plots obtained through the current approach to contour plots obtained from a Green's function approach based on Ref. 19. These contour plots were obtained for the case in which a single plate radiated into the acoustic space and the comparison was done on a qualitative basis.

VI. CONCLUSIONS

This paper describes a mid-frequency technique for modeling structurally excited helicopter cabin interior noise. The model was applied to the Sikorsky S-92 helicopter's interior panel system which has a large cabin enclosed by 43 significant sized panels. Considering the complexity of the system, the results obtained from the simulations were encouraging; the pressure levels in most of the one-third octave bands were found to lie within 3 dB of the experimentally measured values. It must be noted that the accuracy of the results depends strongly on the accuracy of the inputs, which were obtained from a SEA model of the airframe and panel system. SEA has already been shown to be an accurate technique for modeling complex vibrating systems with high modal densities. The hybrid technique combines the strengths of its two individual components to provide results that could not be obtained individually from either technique. The ability of the hybrid model to simulate spatial variations of the interior noise field provides an enabling tool, which could guide the design of active and passive interior noise control systems for rotorcraft.

ACKNOWLEDGMENTS

Technical tasks described in this document include tasks supported with shared funding by the U.S. rotorcraft industry and the Government under the RITA/NASA cooperative agreement number NCCW-0076, Advanced Rotorcraft Technology, August 16, 1995.

[1] C. A. Yoerkie, J. A. Moore, and J. E. Manning, "Development of rotorcraft interior noise control concepts. Phase 1: Definition study," NASA Contractor Report 166101 (1983).

[2] R. H. Lyon and R. G. DeJong, *Theory and Application of Statistical Energy Analysis* (Butterworth-Heinemann, Newton, MA, 1995).

[3] C. A. Yoerkie and P. J. Gintoli, "Helicopter cabin noise—Actual vs statistical energy predictions," *41st Annual Forum of the American Helicopter Society*, Fort Worth, TX (1985).

[4] M. Dussac and A. Morvan, "Evaluation of helicopter internal noise by enhanced statistical energy analysis," *19th European Rotorcraft Forum*, Como, Italy (1993).

[5] J. S. Pollard, "Helicopter gearcase structure-borne noise studies using statistical energy analysis modelling," *1st IME International Conference*, Cambridge, England (1990).

[6] G. P. Mathur, J. E. Manning, and L. R. Quartararo, "Analytical prediction and flight test evaluation of Bell ACAP helicopter cabin noise," *44th Annual Forum of the American Helicopter Society*, Washington, DC (1988).

[7] M. W. Bonilha and D. Young, "Application of predictive and experimental SEA to S-92 interior panel system," United Technologies Research Center Report #6.200.0037-2 (1999).

[8] M. W. Bonilha and F. J. Fahy, "A hybrid model for vibro-acoustic studies," Acust. Acta Acust. **85**, 312–324 (1999).

[9] P. J. Shorter, "Resound: A full spectrum modeling method," *Proceedings of the First International AutoSEA Users Conference*, San Diego, CA (2000).

[10] R. S. Langley and P. Bremner, "A hybrid method for the vibration analysis of complex structural-acoustic systems," J. Acoust. Soc. Am. **105**, 1657–1671 (1999).

[11] J. O'Connell, "Predicting rotorcraft transmission noise," *48th Annual Forum of the American Helicopter Society*, Washington (1992).

[12] A. E. Landmann, H. F. Tillema, and G. R. MacGregor, "Application of analysis techniques for low frequency interior noise and vibration of commercial aircraft," NASA Report number NASA-CR-189555 (1992).

[13] M. A. Wilson and B. L. Josefson, "Combining FEA and SEA in mechanical intensity analysis," AIAA Paper 97-1035.

[14] T. K. Hasselman, M. C. Anderson, and Y-C Lai, "Linking FEA and SEA by principal component analysis," *16th International Modal Analysis Conference*, Santa Barbara, CA (1998).

[15] M. W. Bonilha and F. J. Fahy, "On the vibration field correlation of randomly excited flat plate structures, I: Theory," J. Sound Vib. **214**, 443–467 (1998).

[16] V. V. Bolotin, *Random Vibration of Elastic Systems* (Martinus Nijhoff Publishers, The Hague, 1984).

[17] N. M. Franchek and R. J. Bernhard, "Analytical, numerical, and experimental comparisons of structure-borne noise in a rectangular acoustic enclosure," *Proceedings of Noise-Con 94*, Ft. Lauderdale, FL (1994).

[18] J. Pan and D. A. Bies, "The effect of fluid-structural coupling on sound waves in an enclosure—Theoretical part," J. Acoust. Soc. Am. **87**, 691–707 (1990).

[19] E. H. Dowell, G. F. Gorman III, and D. A. Smith, "Acoustoelasticity: General theory, acoustic natural modes and forced response to sinusoidal excitation, including comparisons with experiments," J. Sound Vib. **52**, 519–542 (1977).

Predicting the vibroacoustic response of satellite equipment panels

S. C. Conlon[a)]
Graduate Program in Acoustics, The Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804

S. A. Hambric
Applied Research Laboratory, The Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804

(Received 14 September 2001; revised 27 September 2002; accepted 2 December 2002)

Modern satellites are constructed of large, lightweight equipment panels that are strongly excited by acoustic pressures during launch. During design, performing vibroacoustic analyses to evaluate and ensure the integrity of the complex electronics mounted on the panels is critical. In this study the attached equipment is explicitly addressed and how its properties affect the panel responses is characterized. FEA and BEA methods are used to derive realistic parameters to input to a SEA hybrid model of a panel with multiple attachments. Specifically, conductance/modal density and radiation efficiency for nonhomogeneous panel structures with and without mass loading are computed. The validity of using the spatially averaged conductance of panels with irregular features for deriving the structure modal density is demonstrated. Maidanik's proposed method of modifying the traditional SEA input power is implemented, illustrating the importance of accounting for system internal couplings when calculating the external input power. The predictions using the SEA hybrid model agree with the measured data trends, and are found to be most sensitive to the assumed dynamic mass ratio (attachments/structure) and the attachment internal loss factor. Additional experimental and analytical investigations are recommended to better characterize dynamic masses, modal densities and loss factors.
[DOI: 10.1121/1.1553462]

PACS numbers: 43.40.Qi [RLW]

I. INTRODUCTION

Over the past decade there has been significant renewed interest in mathematically describing and experimentally evaluating the effects of complex attachments on master structures. There have been many published works describing mostly theoretical research on the subject with applications primarily in the marine and aerospace industries. Examples have been cited by Soize,[1] where the response of a master structure was evaluated computationally considering its attachments as pure mass and its behavior compared to experimental results from the structure with complex ''structural fuzzy'' attachments. The two datasets compared well below about 250 Hz, but from 250 to 3000 Hz the complex structure response peaks were 10–20 dB below the simple predictions, which did not consider the dynamic characteristics of the attachments. As cited by Soize,[1] ''...the rates of dissipation which would correspond to such smoothing are much too high to allow this phenomenon to be explained by mechanical damping alone.'' Indeed, the potentially high levels of master structure-induced damping have recently been explored in numerous works on coupled systems.[2–4] In addition, recent satellite equipment panel damping measurements by Smith[5] made both in air and *in vacuo* clearly illustrate this effect. The results showed, for a single 1.6 kg electronics box mounted on a 650 mm×700 mm sandwich panel, an *in vacuo* loss factor value exceeding 0.03 at 1000 Hz. Smith summarizes: ''The box causes an increase in damping at all frequencies, and particularly at low frequencies where it is increased by a factor of ten.'' Based on Smith's results, one could anticipate that a fully loaded panel with many electronic components could reach levels of *in situ* or induced damping that far exceed this already high (with respect to inherent material damping) level of 0.03.

The effect of the attachments can be viewed as added impedance terms to the master structure primarily contributing both resistive (damping) effects as well as reactive (mass) effects. Previous researchers[6] have found that the (reactive) effect of the equipment is predominantly mass-like or inertial loading rather than local stiffening (although many of the equipment chassis are very stiff, they are attached to the panels via inserts embedded in the panels that have limited stiffness). This was concluded from observing that laser vibrometer obtained displacement contours are only slightly affected by the presence of the equipment. In addition, FEA evaluations modeling the equipment as mass correlated well with measured mode shapes. One could speculate that the added impedances could have significant resulting effects on the master structure's apparent mass, damping, and the total power injected from an external source.

These concepts are illustrated in Fig. 1, which compares individual and spatially averaged acceleration measurements made on two ''identical'' satellite equipment panels: one having actual flight equipment (electronics) mounted on it and the second having the equipment represented with lumped mass (approximating in the test frequency range-

a)Electronic mail: sconlon@sabine.acs.psu.edu

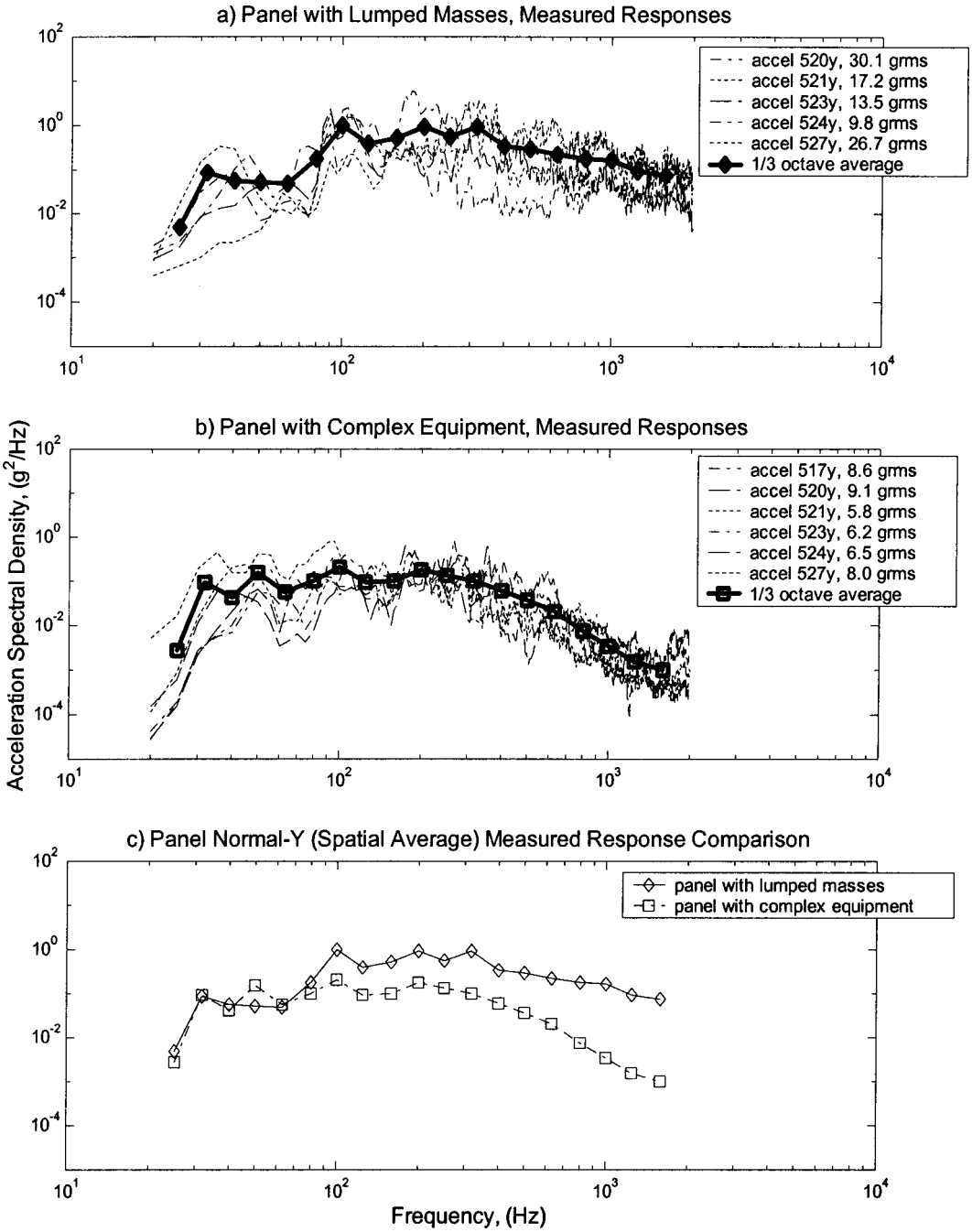


FIG. 1. A comparison of response measurements: (a) panel with lumped masses, (b) panel with complex equipment, (c) a comparison of spatial averages.

nonresonant/very low modal density attachments). The panels are portions of one side of a large enclosed box-like satellite structure with the equipment mounted on the interior. Note that the subsequent analysis simulates this by assuming the panel to be simply supported and baffled with one-sided excitation (on the side without the equipment.) The response of these types of panels tends to be dominated by the direct acoustic excitation and the equipment response has been shown to be dominated by the panel vibration. This is true even when the equipment side of the panel is directly exposed to the external acoustic field.[6] The equipment loaded test panel is shown in Fig. 2. Both test panels were excited by reverberant acoustic fields, with the response measurements shown normalized to the same input amplitude. As can be seen in Fig. 1(c), the two response spatial averages are nearly the same at low frequencies (below 100 Hz), where both sets of attachments are contributing mainly in a reactive sense or as a lumped mass addition. The mid-frequency range ($\sim$100–400 Hz) shows varying degrees of difference between the measurements, where the flight equipment begins to resonate and the lumped mass simulators are still approximately ''nonresonant.'' At higher frequencies ($>$400 Hz) the two measurements rapidly diverge with that of the real equipment loaded panel being approximately 20 dB below that of the lumped mass loaded panel at 2000 Hz.

The data shown illustrates the significant effects on the panel responses due to the vibration characteristics of the attachments. Typical analytical modeling techniques would

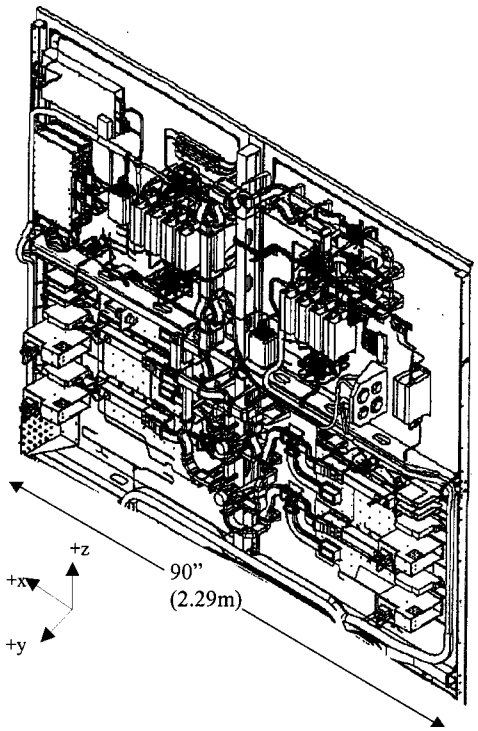


FIG. 2. Complex equipment-loaded panel layout.

not address or account for these measured differences in response explicitly, since the equipment or attachments are often accounted for via estimations or ''guesses'' as to their induced damping effects and added mass to the master structure. Based on the measured data it is clear that more insight and better predictive/modeling techniques are required to explicitly account for how the complex equipment affects the master structure to which it is attached. In addition, explicit modeling of the attachments will allow the estimation and evaluation of their responses, which may be of utmost importance and can be quite different than the master structure response.

Many methods have been used for developing Statistical Energy Analysis (SEA) predictions of equipment-loaded panel responses. To model coupling between panels and acoustic spaces, the aerospace industry typically applies some form of panel radiation efficiency formulation based on work by Maidanik[7] on rib-stiffened (unloaded) panels. The equipment, which often has a total mass many times greater than the panel mass, is typically either smeared into the panel as limp mass and/or used for scaling down an unloaded structure response prediction. Thin plate formulations are often used for the panel wave properties and considering the typical sandwich panel designs used for space applications, this often leads to estimates that vary significantly from those considering panel transverse shear effects. The predictions based on these methods are often inconsistent with measured data, usually due to the mathematical models not encompassing the significant features of the problem at hand.

A large compendium of useful information related to the vibroacoustic characterization of satellite responses is presented in the European Space Agencies Structural Acoustics Design Manual.[8] The manual discusses the need for accounting for equipment-loading effects, but no clear systematic approach is presented. However, the manual does indicate that this area is not yet ''fully researched.''

The overall objective of this study is to develop improved procedures for the modeling and analysis of lightweight equipment-loaded panels. A critical aspect of the study will focus on explicitly addressing the panel nonhomogeneous features and attachments and developing methods that will predict the significant differences observed between the equipment versus mass loaded panels shown in Fig. 1 (20 dB at high frequencies). Industry standard methods do not explicitly address the nature of the attachments and thus cannot predict the significant differences shown in Fig. 1. In addition, the standard methods use generalized procedures for estimating the vibration characteristics of the panels that do not account for the unique nonhomogeneous nature of most satellite equipment panels. Current methods are examined and improved upon utilizing the specific test data presented and through additional computational investigation with Finite Element Analysis (FEA) and Boundary Element Analysis (BEA) methods.

II. SEA MODEL FOR PANEL WITH MULTIPLE ATTACHMENTS

A. Model description

Figure 2 shows a typical communications satellite equipment panel and is the specific design to be investigated. The basic structure is a sandwich/honeycomb core flat panel with various stiffeners. The attached equipment are complex electronics and associated hardware. These types of equipment are mounted to the panel via inserts embedded in the panel core. The insert core interface is of limited stiffness, thus the reactive effect of the equipment on the master panel is expected to be predominantly mass-like. The measured panel transverse vibration data, from reverberant chamber acoustic testing are acceleration spectral density spectra of two basic types:

(i) panel responses for the panel with ''nonresonant'' lumped mass attachments (replacing the electronic equipment overall mass distribution), and
(ii) panel responses for the panel with complex equipment attachments (as shown in Fig. 2).

The actual mass loadings of these two panels are compared in Table I with the structural properties. (Note the total lumped mass loading is somewhat greater than the equipment mass loading in total mass.) Additional panel structural detail is given in Fig. 3. Clearly, the panel assembly is a lightweight stiff structure with many complex attachments to be considered.

In this section a new SEA model is established that will serve as the baseline utilized for all subsequent refinements presented in this investigation. In the new model, the resonant attachment modes are addressed explicitly and each represented as single degree of freedom systems coupled to a finite plate. The associated nonresonant mass of the attachments now becomes part of the panel ''effective mass,'' which this new model framework also explicitly addresses. The new model is then exercised to help identify modeling

TABLE I. Panel properties.

Property	Loaded panel			
	Value	Units (In. lb. s)	Value	Units (m kg s)
Panel face sheets				
Material (Al alloy)				
Thickness	0.01	in.	2.540E−04	m
E, modulus of elasticity	1.05E+07	psi	7.239E+10	P (N/m^2)
Density	0.101	$lb./in.^3$	2.796E+03	kg/m^3
Poisson's ratio	0.33	···	0.33	···
Panel core				
Material (Al alloy)				
Thickness	0.75	in.	1.905E−02	m
G (ribbon, shear modulus)	27,000	psi	1.862E+08	P
G (warp, shear modulus)	13,000	psi	8.963E+07	P
Density	1.157E−03	$lb./in.^3$	3.204E+01	kg/m^3
Panel width/height dimensions				
Width	90.0	in.	2.286E+00	m
Height	72.5	in.	1.842E+00	m
Panel face sheets and core (weight or mass)	18.84	lbs.	8.546E+00	kg
Misc. struct. (stiffeners, heat pipes, corefill, etc.)	92.00	lbs.	4.173E+01	kg
Tot. structure	110.84	lbs.	5.028E+01	kg
w/complex equipment (estimated values)				
Equipment & Misc. (weight or mass)	106.62	lbs.	4.836E+01	kg
Total (Loaded Panel) weight or mass	217.5	lbs.	9.864E+01	kg
w/"non-resonant" attachments (estimated values)				
Lumped Masses & Misc. (weight or mass)	131.87	lbs.	5.981E+01	kg
Total (loaded panel) weight or mass	242.7	lbs.	1.101E+02	kg

parameters that require a more detailed investigation. This model will serve as a backbone for incorporating the subsequent detailed parameter investigation results. The basic approach focuses on adding modeling flexibility and allowing response predictions for both panel and attachments.

The equipment-loaded panel is schematically represented in Fig. 4, with the subsystem parameters defined as $\pi_{ea}(\omega)$, the acoustic volume external power input; $E_1(\omega)$, $E_2(\omega)$, $E_3(\omega)$, the subsystem temporally and spatially averaged total energies; $n_1(\omega)$, $n_2(\omega)$, $n_3(\omega)$, the subsystem modal densities; and $\eta_{ij}(\omega)$, for $i=j$ the subsystem damping loss factors, for $i\neq j$, the coupling loss factor from subsystem i to subsystem j.

Note, π_{ea}, η, n, and E are all functions of frequency ω. Frequency dependence is implied for these quantities throughout the analysis.

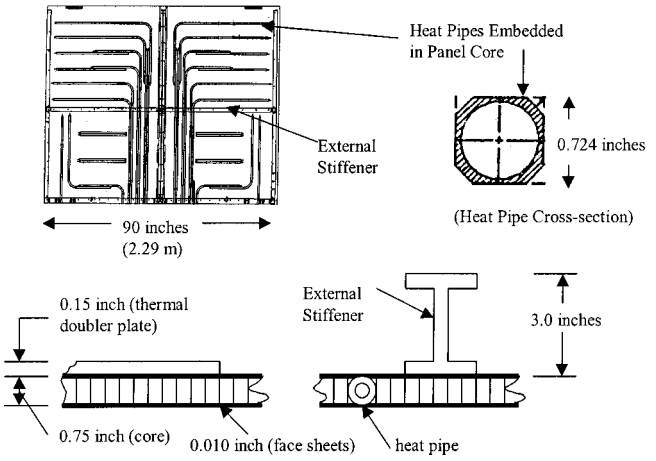


FIG. 3. Panel structure details.

B. Acoustic volume response energy and modal density

The temporally averaged energy of the acoustic volume is specified/known through its relationship to the mean square spatial average pressure:[9]

$$\langle p_1^2\rangle = \frac{\rho_0 C_0^2 E_1}{V_1}, \tag{1}$$

where $\langle p_1^2\rangle$=measured acoustic volume-averaged mean square pressure given in Table II: V_1=acoustic system volume, ρ_0=density of acoustic medium, and C_0=the speed of sound in an acoustic medium.

The acoustic volume modal density can be approximated as[9] (considering the volume term only)

$$n_1 = \frac{\omega^2 V_1}{2\pi^2 C_0^3}. \tag{2}$$

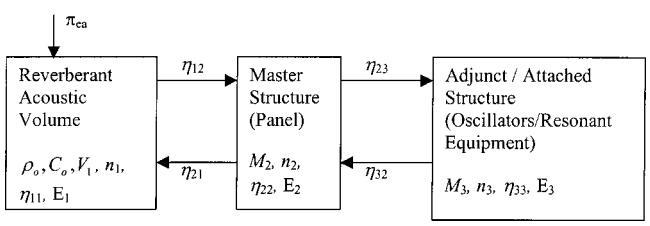


FIG. 4. SEA system model of panel with multiple complex equipment attachments.

TABLE II. Reverberant test chamber acoustic pressures.

1/3 octave band center frequency (Hz)	Acoustic chamber Average of four microphones SPL (dB *re* 20 μPa)
25	121.9
31.5	125.3
40	126.6
50	129.5
63	130.8
80	134.3
100	136.9
125	136.2
160	136.9
200	137.8
250	139.2
315	139
400	136.8
500	133.4
630	132.6
800	130.5
1000	129
1250	127.8
1600	126.5
2000	125.2
2500	123.7
3150	122.5
4000	121.1
5000	120.1
6300	119.4
8000	118.5
10000	116.2
Overall SPL	147.4

C. Complex panel and equipment mass and modal densities

The partitioning of the equipment and panel dynamic or effective mass over frequency is uncertain since the panel subsystem has an added mass that is based on the nonresonant portion of the attachments. The added equipment mass alters the vibration properties of the panel. At low frequencies (below any of the attached equipment resonances) the total mass of the panel and equipment will be associated with the modes of the ''built-up'' panel. There will be some transition region through the mid-frequencies with some potential limit reached at higher frequencies, as more equipment mass becomes associated with equipment resonances. The loaded panel modal density is calculated using the panel mass/area as the effective mass including the panel structure and some fraction of the attachment mass.

The sandwich panel transverse or effective bending wave speed is[10] (which accounts for transverse shear effects):

$$C_{B\,\mathrm{eff}}=\left[\frac{1}{C_B^3}+\frac{1}{C_S^3}\right]^{-1/3}. \tag{3}$$

This approximation from Rindell is based on earlier work by Kurtze and Watters,[11] but does not encompass the high-frequency region of single-face sheet bending-dominated waves. The bending and shear waves speeds in Eq. (3) are

$$C_B=\left[\frac{D}{\mu_S}\right]^{1/4}\sqrt{\omega} \tag{4}$$

and

$$C_S=\left[\frac{GH}{\mu_S}\right]^{1/2}, \tag{5}$$

where $D=Et_{fS}(H+t_{fS})^2/2(1-\nu^2)$, the section bending stiffness, E=the face sheet modulus of elasticity, t_{fS}=the face sheet thickness (single face sheet), H=the core thickness or depth, ν=the face sheet Poisson's ratio, $\mu_S=2\rho_{fS}t_{fS}+H\rho_c+(\textit{nonresonant equipment})$, the built-up panel mass/area, ρ_{fS}=the face sheet density, ρ_c=the core density, and G=the core shear modulus$=\sqrt{G_RG_W}$, the geometric mean of the ribbon and warp direction values.

The bending wave speed is dispersive and the shear wave speed is not. In addition, when the bending wavelengths are large compared to the thickness; then $C_{B\,\mathrm{eff}}$ reduces to thin plate theory giving $C_{B\,\mathrm{eff}}\sim C_B$.

The panel effective modal density can be written as[10] (where A_2 is the panel surface area),

$$n_2=\frac{A_2\omega}{2\pi C_{B\,\mathrm{eff}}^2}\left(1-\frac{\left[\frac{1}{C_B^3}+\frac{1}{C_S^3}\right]^{-4/3}}{2C_{B\,\mathrm{eff}}C_B^3}\right). \tag{6}$$

The attached resonant equipment is estimated to have a total constant modal density of approximately 0.043 modes/(rad/s) that comes from an estimated mode count of 5.0 modes in the 80 Hz one-third octave band. This modal density estimate is based on the attached equipment design criteria (most of the equipment is designed to have no resonances below 75 Hz; some are higher) and the assumption that the equipment modes are predominantly plate-like. The sensitivity of the overall response prediction with respect to the assumed resonant equipment modal density will be evaluated later in the investigation.

D. Damping loss factors and loss factor definitions

The built-up panel structure damping loss factor is estimated to be

$$\eta_{22}=0.050 \text{ up to } 500 \text{ Hz};$$

$$\text{then } 0.050\left(\frac{500}{f}\right)^{1/2} \text{ above } 500 \text{ Hz}.$$

The attached resonant equipment average damping loss factor is estimated to be approximately 0.066 (0.033 percent critical damping). This average value is based upon measurements made on similar types of equipment both *in situ* (on the panels) and isolated (in a fixed base configuration on a shaker table).

The definitions and notations of the various measures of loss factor used in the analysis are now reviewed. Definitions and notations adopted follow recent works by Maidanik *et al.*[12–16] The panel and equipment subsystems are shown for an uncoupled system (panel only) and coupled system (panel and attachments) in Fig. 5. The ''^{0}'' superscript indicates power input and response energy levels for the un-

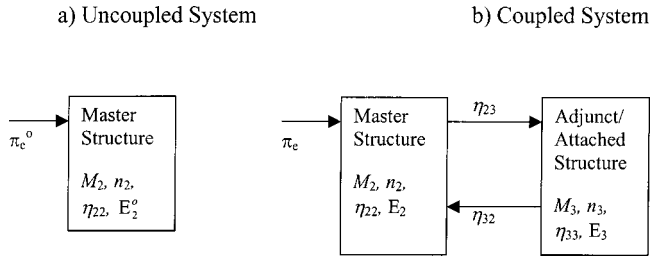


FIG. 5. Uncoupled and coupled SEA system models: (a) uncoupled system, (b) coupled system.

coupled system. This same uncoupled system input power is also traditionally used in SEA for the coupled system input power, which ignores the effects of the attachments. This work will show that the power input and resulting energy levels for equipment-loaded lightweight panels must consider the effects of the attachments when calculating the external (panel) input power.

The master structure (panel) damping loss factor in isolation is

$$\eta_{22}=\frac{\pi_e^0}{\omega E_2^0}, \tag{7}$$

where π_e^0 and E_2^0 are the master structure uncoupled input power and total energy. The adjunct structure (attached equipment) damping loss factor in isolation is similarly defined and noted simply as η_{33}.

The apparent added damping to the master structure (relating the *in situ* effects of the adjunct structure on the master structure with respect to the external input power) for the coupled systems of Fig. 5 is

$$\eta_s=\eta_{32}\frac{n_3}{n_2}\left\{1-\frac{\eta_{32}}{\eta_{33}+\eta_{32}}\right\}. \tag{8}$$

The induced damping of the master structure *in situ* is

$$\eta_t=\eta_0+\eta_s=\eta_{22}+\eta_s=\frac{\pi_e}{\omega E_2}, \tag{9}$$

where π_e and E_2 are the master structure coupled input power and total energy. *Note*: Both η_s and η_t are meaningful ratios of dimensionless dynamic quantities and can take on values significantly greater than unity, however, they are not true loss factors.[16] The true effective loss factor of the system relates the *in situ* external input power to the total system energy:

$$\eta_e=\frac{\pi_e}{\omega(E_2+E_3)}. \tag{10}$$

Unlike the induced loss factor η_t, the effective loss factor η_e is bounded with asymptotic limits of the "*in isolation*" loss factors of the master structure (η_{22}) and the adjunct structure (η_{33}). These defined loss factors will be contrasted and compared later in the investigation of specific subsystem parameters as applied to equipment-loaded panels.

E. Resonant equipment oscillator model

The model chosen for the resonant equipment, the conceptually simplest available, is that of a simple single degree of freedom system applied to each equipment mode of vibration. Using this description, the coupling from a single oscillator to the panel is given by Lyon[9] as

$$\eta_{\text{coup}}=\omega_0 M\langle G_p\rangle, \tag{11}$$

where $\langle G_p\rangle$ is the average panel conductance. Applying Eq. (11) to a distribution of oscillators (resonant equipment modes) leads to the average equipment to panel coupling loss factor,

$$\eta_{32}=\omega\frac{M_3}{N_3}\langle G_2\rangle, \tag{12}$$

where M_3 is the total resonant equipment mass, N_3 is the total mode count in the analysis band for the resonant equipment, and $\langle G_2\rangle=(\pi/2)(n_2/M_2)$ is the average panel conductance.

F. Response analysis

The complex equipment-loaded panel is now modeled using SEA (Fig. 4). The panel structure is coupled to a distribution of oscillators representing the resonant equipment. The acoustic volume pressure is specified leaving two SEA power balance equations[9] describing the panel and equipment responses:

$$\omega\eta_{12}E_1+\omega\eta_{32}E_3-\omega\eta_{22}E_2-\omega\eta_{21}E_2-\omega\eta_{23}E_2=0, \tag{13}$$

$$\omega\eta_{23}E_2-\omega\eta_{33}E_3-\omega\eta_{32}E_3=0. \tag{14}$$

The panel-induced damping now becomes

$$\eta_t=\eta_{22}+\frac{\omega M_3}{\Delta n_2}\langle G_2\rangle\left[1-\left\{1\Big/\left(\frac{\eta_{33}\Delta n_3}{\omega M_3\langle G_2\rangle}+1\right)\right\}\right]. \tag{15}$$

The SEA power balance equations can now be solved to obtain the subsystem responses.

G. Prediction results

The formulations developed for a panel with multiple attachments are now applied to the equipment-loaded panel. For now, the resonant fraction of equipment mass is assumed to be 1.0 for all frequencies (nonresonant equipment mass would be associated with the panel subsystem). The equipment panel and acoustic volume properties are as given in Tables I and III. The panel and equipment responses are predicted with the input acoustic volume pressures given in Table II. The total structure mass is used here for the panel mass, that is, the miscellaneous structure mass from Table I has been evenly distributed on the panel as a limp mass. The added mass affects the panel radiation efficiency, for now computed via a generalized form[4] of the NASA Lewis

TABLE III. Acoustic volume properties.

Property	Acoustic volume			
	Value	Units (in. lb. s)	Value	Units (m kg s)
Acoustic volume				
Media (air @20 °C)				
Volume (55 000 ft^3 test chamber)	9.504E+07	$in.^3$	1.557E+03	m^3
Density	4.371E−05	$lb./in.^3$	1.210E+00	kg/m^3
Speed of sound	1.350E+04	in./s	3.429E+02	m/s

method,[17] by reducing the wave speed (also increasing the modal density) as well as raising the panel critical frequency. The higher critical frequency results in a lower panel radiation efficiency over much of the frequency band of interest. Figure 6 shows the various system loss factors. In Fig. 6(a) the oscillator damping loss factor is compared to the oscillator to panel coupling loss factor. The coupling loss factor greatly exceeds the damping loss factor, which indicates strong coupling where the oscillator and panel modal energies are nearly equal $E_3/n_3 \cong E_2/n_2$ (the equipartition of modal energy). Figure 6(b) compares the subsystem loss factors. Although the induced panel loss factor reaches a level greater than 0.2, the true system effective loss factor is bounded by the in isolation panel and oscillator levels.

The panel and oscillator predicted average responses are compared with the measured data in Fig. 7. Although the panel and oscillator predictions are somewhat representative of the measured data, significantly better correlation is desired, especially with respect to the panel average. Less credence is given the measured equipment response since it is a single measurement and is included only for rough comparisons. Various iterations of the assumed parameter values (i.e., oscillator/panel mass and modal density ratios, panel radiation efficiency, damping loss factors, etc.) show the response to be dependent (to varying degrees) on these values. In addition, the SEA model does not address the nonhomogeneous stiffness and mass features of the equipment-loaded

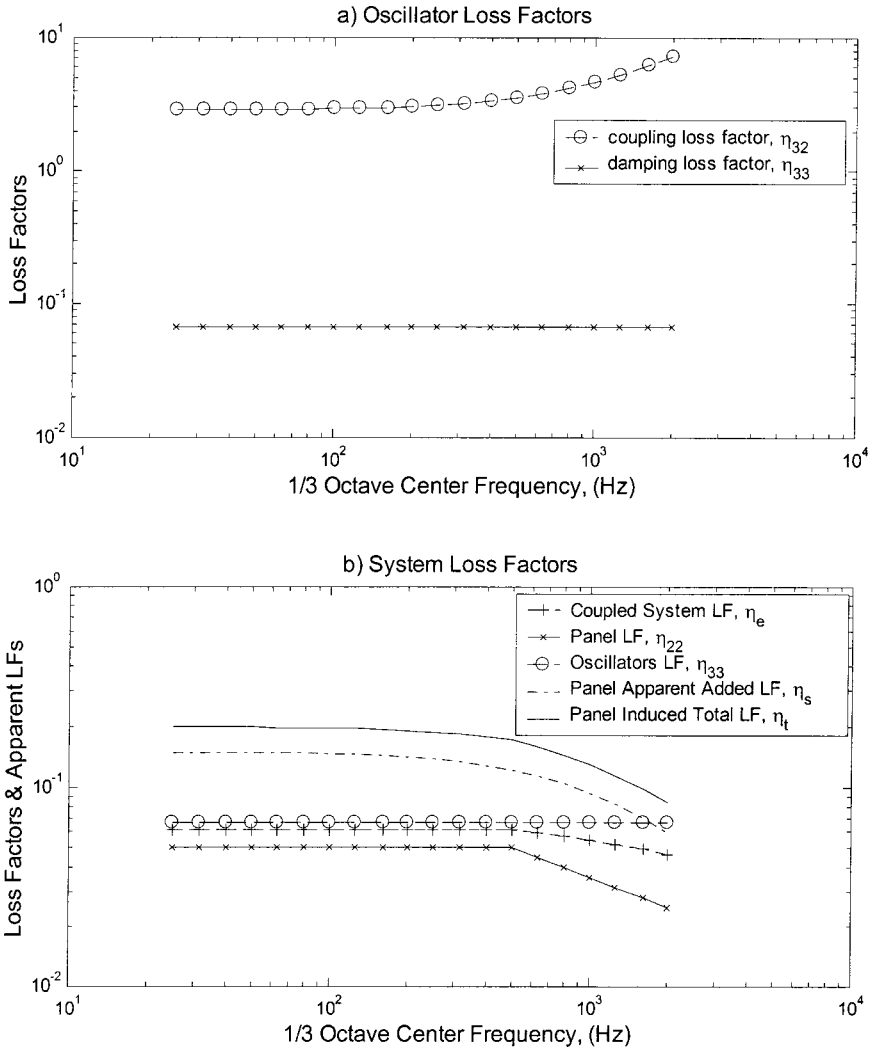
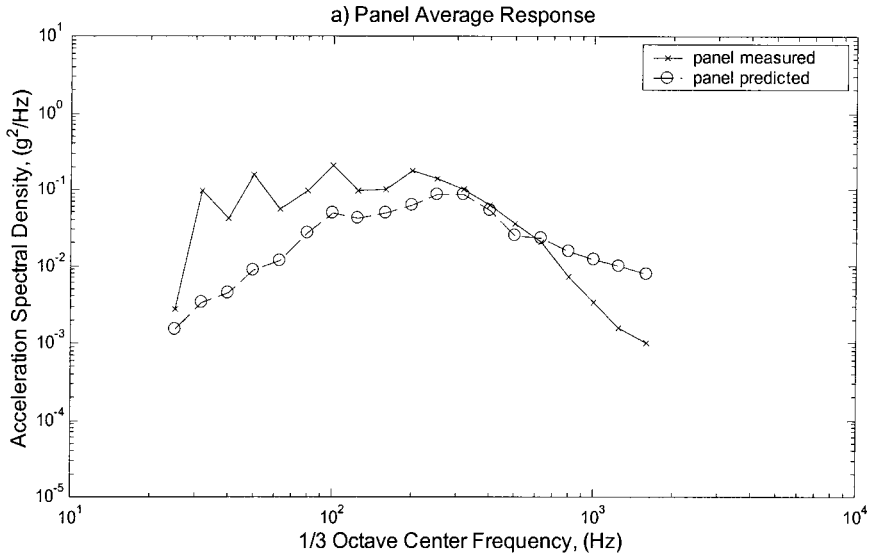


FIG. 6. Loss factors: (a) oscillator coupling and damping loss factors, (b) a comparison of system, panel, oscillator, panel apparent, and panel-induced loss factors.



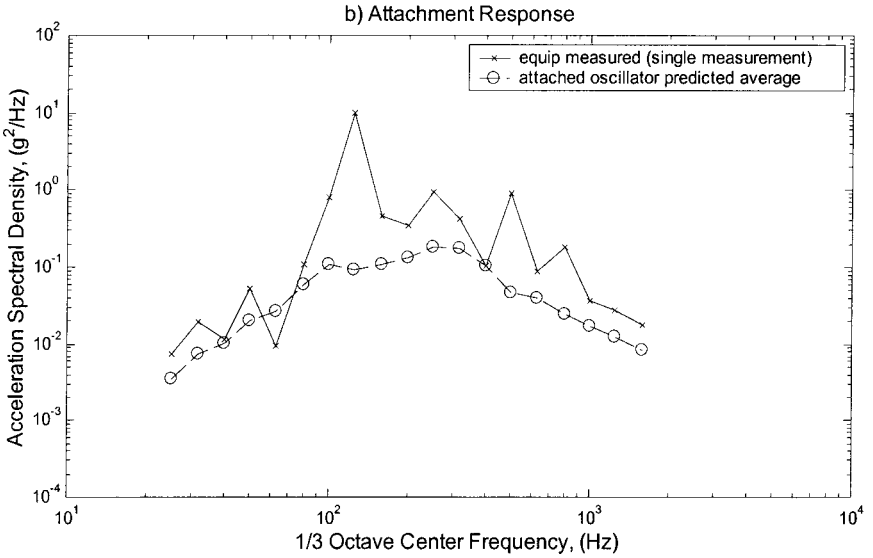


FIG. 7. Predicted versus measured responses: (a) panel spatial average, and (b) equipment.

panel, and thus any of their potential effects on the critical parameters listed.

H. Parameters for detailed investigation

Many of the parameters identified in the previous sections require further investigation in order to enhance the accuracy of the complex structure response predictions. A parametric response sensitivity analysis was performed previously for equipment-loaded lightweight panels.[2] The sensitivity analysis began by first modifying the standard SEA power input based on recent work by Maidanik and Dickey,[12] to account for internal system couplings. The coupled system response parameter sensitivities were then assessed. The key results of the modified input power parameter sensitivity analysis were as follows.

(i) Internal couplings of strongly coupled systems can significantly affect the external input power.
(ii) The panel response is very sensitive to the effective dynamic mass ratio of the equipment/panel (attachments/master structure) and the equipment internal damping.
(iii) With strongly coupled equipment and an equipment/panel modal density ratio greater than unity, the response prediction is weakly sensitive to the panel internal damping and equipment/panel modal density ratio.

The modified SEA power input results[2] are presented and applied in Sec. IV of this paper. The nonhomogeneous nature of the panel structure can also significantly affect its radiation coupling, equipment coupling and modal density. These specific parameters are investigated in Sec. III of this paper.

III. REVISED MODELING OF HARDWARE

A. Background on hybrid SEA

The SEA predictions developed in this section are often referred to as ''Hybrid SEA'' applications in that they incor-

TABLE IV. FEA panel configurations.

No.	Configuration title	Comments	Total model Weight (lbs.)/Mass (kg)
1	Unloaded	Panel face sheets and core only	18.84/8.546
2	Total panel structure	Panel face sheets/core, external stiffeners, center strut, internal heat pipes, thermal doubler plates, adhesive, core fill, inserts, etc.	110.8/50.28
3	Total panel structure +equipment mass	Total panel structure with equipment mass (as limp mass) distributed locally	217.5/98.64
4	Total panel structure +lumped masses	Total panel structure with lumped masses in appropriate locations	242.7/110.1

porate the formulations of traditional SEA as well as deterministic results from FEA and BEA. When applied to the analysis of complex structural acoustic systems, SEA is often associated with high frequencies and FEA with low frequencies. This segregation often results from balancing practicalities in implementation versus desired confidence in the predictions. Another product of the typical implementation of SEA and FEA is what is commonly referred to as ''The Mid-Frequency Gap''—the overlapping frequency region, where it is often believed that confidence in both methods separately is low. In recent years several mid-frequency and hybrid or combination approaches have been proposed. Of note are works by Soize,[1] Lu,[18] Langley and Bremner,[19] Shorter,[20] and Manning.[21,22] The readily tractable approach taken in this chapter using the complex panel vibration characteristics from detailed finite element analyses is analogous to the approach taken by Manning[22] and the utilization of the BEA modal radiation efficiency results also draws from previous work by Manning.[21] In essence, using the FE/BE results to enhance the SEA analysis puts information and detail back

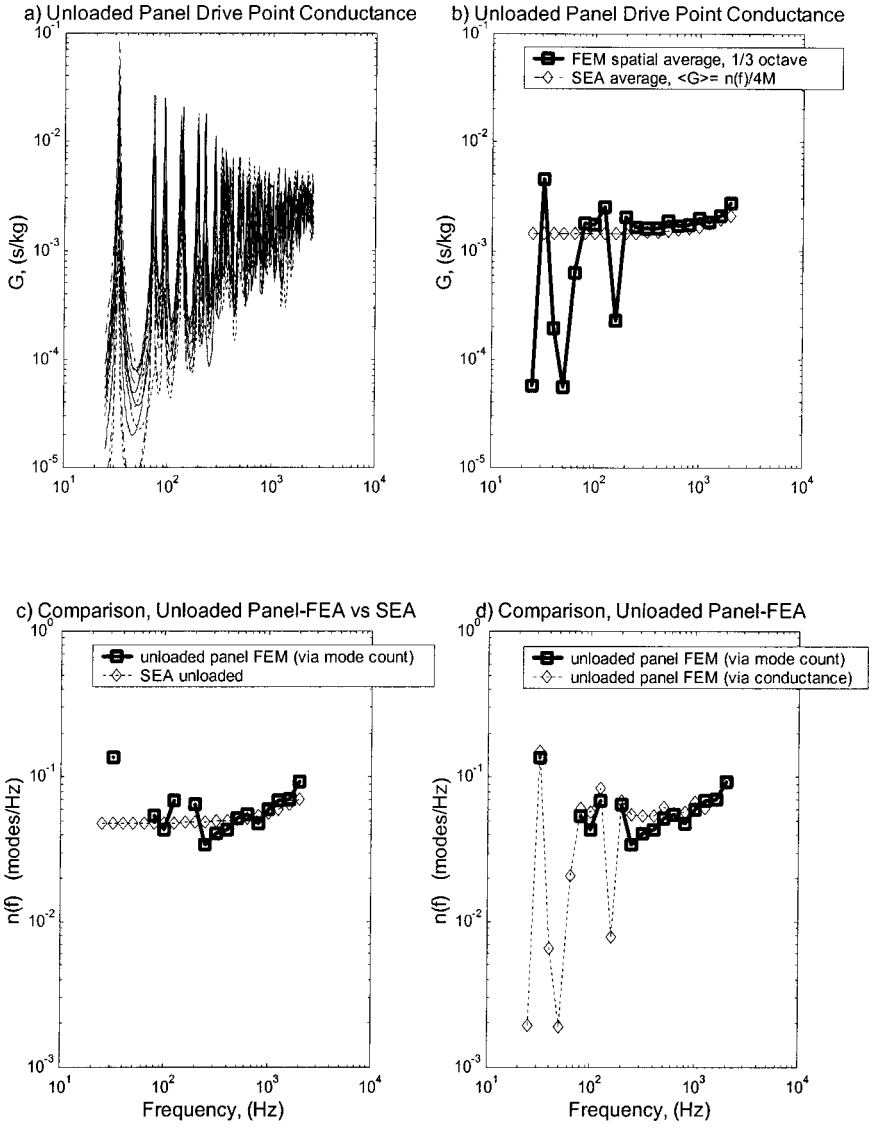


FIG. 8. Unloaded panel FEA results: (a) unloaded panel FEA drive point conductance, (b) unloaded panel average FEA versus SEA conductance, (c) unloaded panel FEA (via mode count) versus SEA modal densities, (d) unloaded panel FEA (via conductance) versus SEA modal densities.

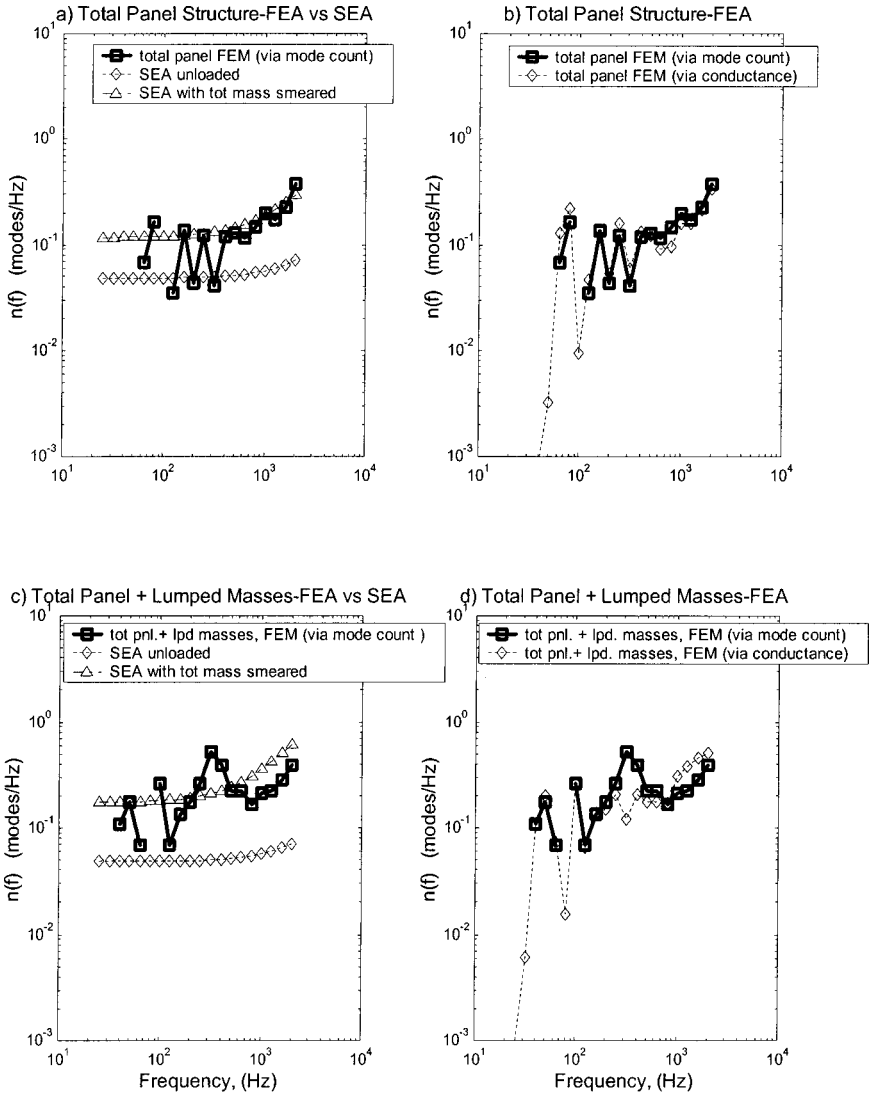


FIG. 9. Total panel FEA results: (a) total panel structure FEA (via mode count) versus SEA modal densities, (b) total panel structure FEA (via conductance) versus SEA modal densities, (c) total panel+lumped masses structure FEA (via mode count) versus SEA modal densities, (d) total panel +lumped masses structure FEA (via conductance) versus SEA modal densities.

into the analysis that was inherently sacrificed by using the SEA method. Once the FE/BE results are obtained, they will be used in a revised SEA model with the goal of better predicting the response trends of equipment and lumped mass-loaded panels similar to those shown in Fig. 1.

B. Panel finite element analyses

In this section the commercial finite element code MSC/NASTRAN[23] is used to analyze the detailed models that accurately characterize the nonhomogeneous panel vibration characteristics. The baseline panel (Fig. 3, 90 in.×72.5 in.) is modeled using approximately 1870 QUAD4 elements with the properties from Table I to properly characterize the sandwich panel bending and shear characteristics. QUAD elements are also used to model the thermal doublers and BAR elements are used to represent the beam-like features, as shown in Fig. 3. The finite element analyses are used to compute the panel spatial average conductance and modal density for the four configurations listed in Table IV, which are then compared with asymptotic results obtainable through SEA approaches. The FEA results are obtained by performing modal frequency response analyses (SOL111[23]) using unit input point forces, thus obtaining the panel drive point mobilities. Although the response analyses are only performed up to 2500 Hz, the panel modes up to 3500 Hz are included to ensure adequate convergence of the modal summation solution (based on convergence studies performed). Twelve "randomly" distributed points are analyzed, assuming the panel perimeter to have pinned boundary conditions. These same drive points and the baseline panel finite element mesh are used for all the configurations analyzed. *Note*: previous iterative studies varying the damping used in the FEA response analyses have shown that the computed spatially averaged one third octave band conductance is relatively insensitive to the assumed loss factor for typical values of approximately 0.01–0.05. The panel average conductance $\langle G \rangle$ is estimated by spatially averaging the real part of the drive point mobilities. The panel modal density is calculated in two ways for comparison (which is of special interest for the configurations with nonhomogeneous features). The first method is by counting the modes per band from the finite element analysis,

$$n(f) = \frac{N_{\Delta f}}{\Delta f}. \tag{16}$$

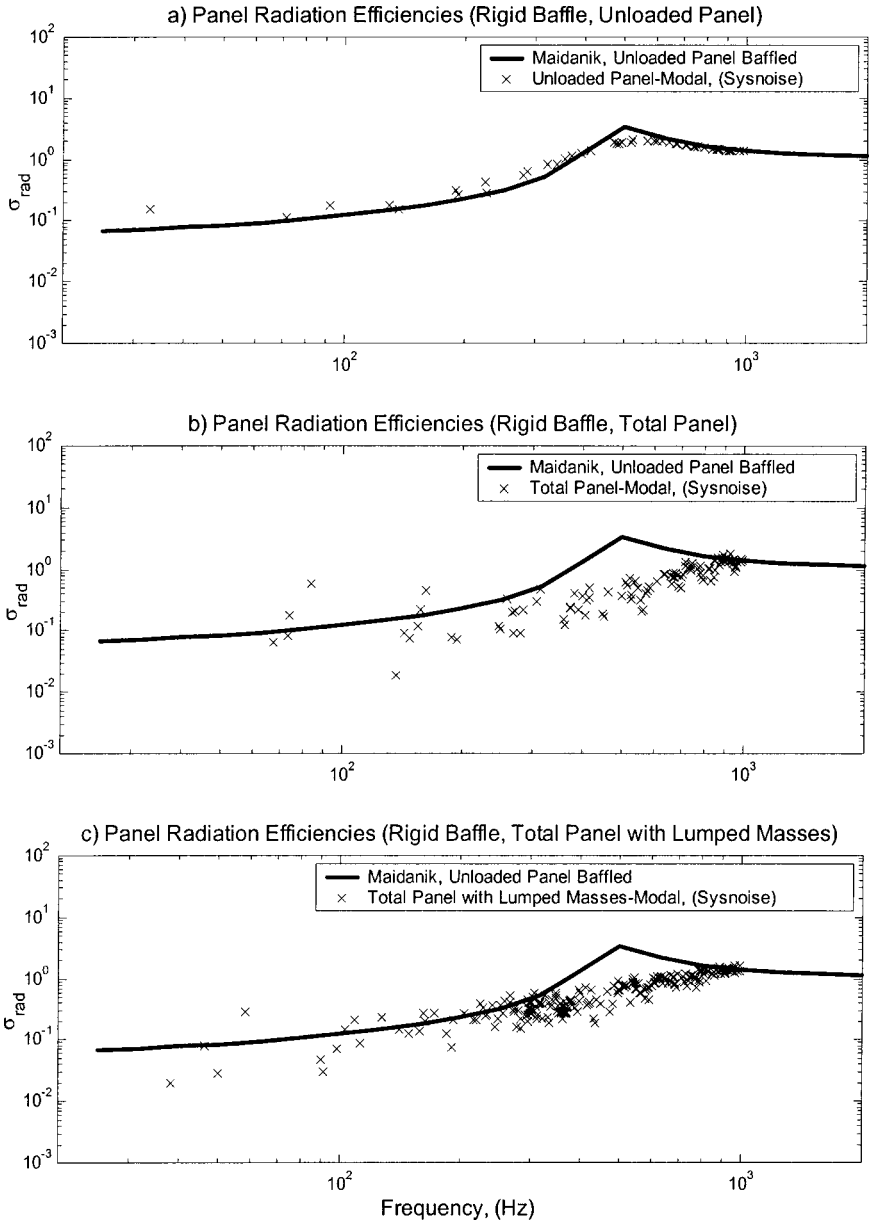


FIG. 10. BEA results: (a) unloaded panel BEA computed modal versus average theoretical radiation efficiency, (b) total panel structure BEA computed modal radiation efficiencies versus unloaded panel average theoretical radiation efficiency, (c) total panel structure+lumped masses BEA computed modal radiation efficiencies versus unloaded panel average theoretical radiation efficiency.

The second method is through the modal density relationship with the average conductance,

$$n(f)=4M\langle G\rangle. \tag{17}$$

Equation (17) has been derived by Cremer *et al.*[24] for a homogeneous plate, i.e., with constant surface mass density. Norton[25] has suggested that Eq. (17) is also applicable to nonhomogeneous plates, where the mass is the total mass of the structural elements, although this is a topic currently under debate in the SEA community. In this work the mass in Eq. (17) is assumed to be the total mass for the uniform as well as nonhomogeneous structures. Figure 8(a) shows the FE computed conductances for the 12 random locations on the panel. Figure 8(b) shows the good comparison for the uniform panel between the FE spatial and frequency average compared to the typical SEA value (note that both formulations include transverse shear effects that become significant above approximately 1000 Hz). Figures 8(c)–9(d) compare the FE and typical SEA results for configurations 1, 2, and 4 from Table IV.

C. Boundary element analyses

In this section the panel radiation characteristics are investigated. In a SEA analysis the panel radiation efficiency σ_{rad} is typically used to calculate the panel to acoustic volume coupling via the relation[9]

$$\eta_{\text{rad}}=\frac{\rho_0 C_0 A_p}{\omega M_p}\sigma_{\text{rad}}, \tag{18}$$

where A_p and M_p are the panel area and mass and $\rho_0 C_0$ is the acoustic medium characteristic impedance (and the acoustic volume to panel coupling is calculated using η_{rad} along with the consistency relationship). The radiation efficiency is thus a critical parameter in the response prediction since it is indicative of the power input to the panel system. The finite element analysis determined panel mode shapes

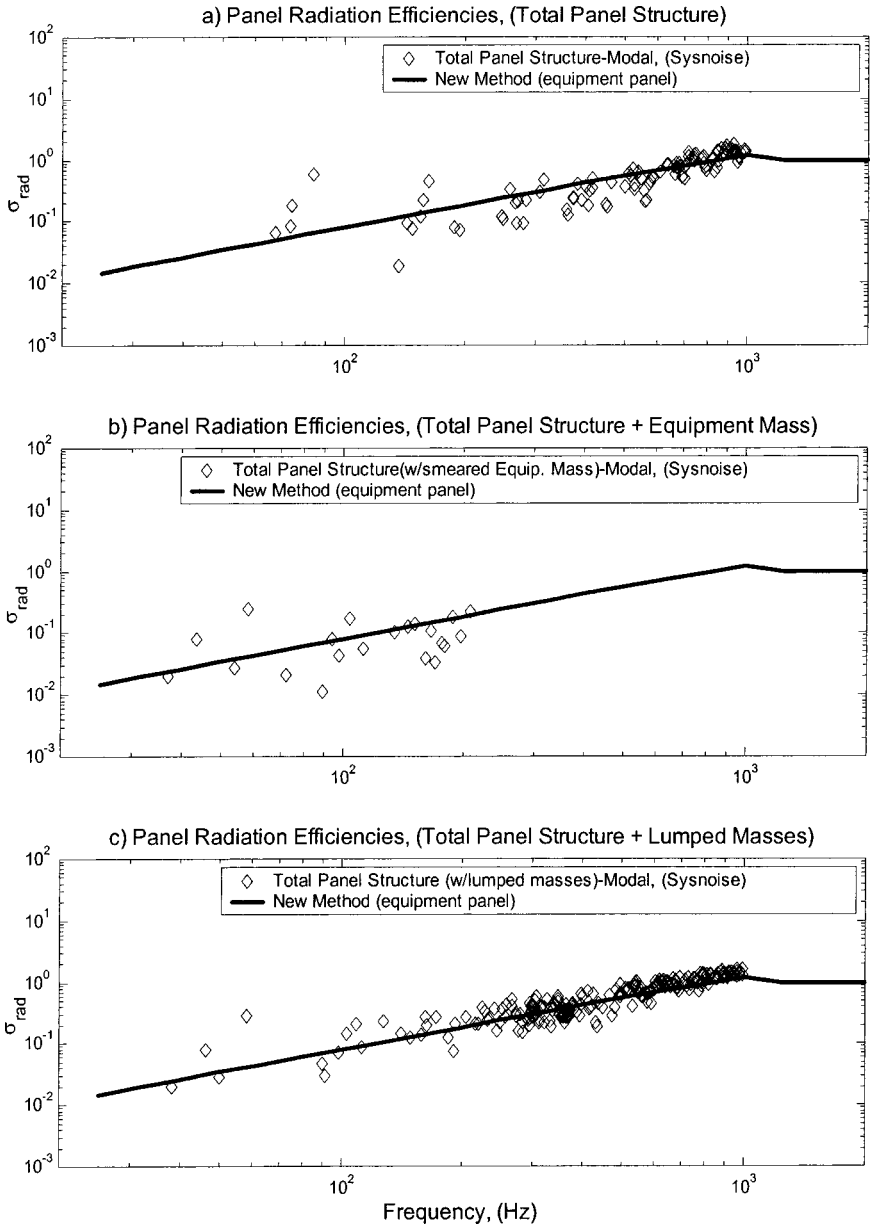


FIG. 11. Equipment panel BEA computed radiation efficiencies versus curve fit from Eq. (40): (a) total panel structure, (b) total panel structure +equipment mass, (c) total panel structure+lumped masses.

from the previous section are input to the commercial vibroacoustic analysis code SYSNOISE[26] to compute their radiation efficiencies. The panel is assumed baffled and the FEA mode shapes are used to define a velocity boundary condition for BEA in SYSNOISE using the acoustic properties from Table III. The calculation results from the BEA are of the input power (π_i), output power (π_0) and the radiation efficiency (σ_{rad}) and are given as[26]

$$\pi_i = \rho_0 C_0 \oint_S |\nu_{i-\text{rms}}^2| dS, \tag{19}$$

$$\pi_0 = (1/2) \oint_S p \cdot \nu^* \, dS, \tag{20}$$

and

$$\sigma_{\text{rad}} = \frac{\pi_0}{\pi_i}, \tag{21}$$

where ν are the defined panel surface normal velocities (input), p are the solved for surface pressures (computed by SYSNOISE), S is the panel radiating surface area, and ρ_0 and C_0 are the acoustic medium density and sound speed.

The same approach as was taken in the finite element analyses is repeated here when computing the panel modal radiation efficiencies, starting with the simple spatially homogeneous panel and adding in increments the nonhomogeneous complex features. These calculations are made using the mode shapes from the configurations in Table IV. For comparison purposes, calculations are made for Maidanik's[7] well-known result for baffled unloaded panels. Figure 10(a) shows the good comparison between the BE calculated (baffled) unloaded panel radiation efficiency and Maidanik's result, with the peak occurring at the critical frequency. Figures 10(b) and (c) compare the total nonhomogeneous panel (111 lbs.) and the total panel with lumped masses (243 lbs.). The BEA modal results sets are used to develop approximate curve fits for a subsequent analysis. The computed modal

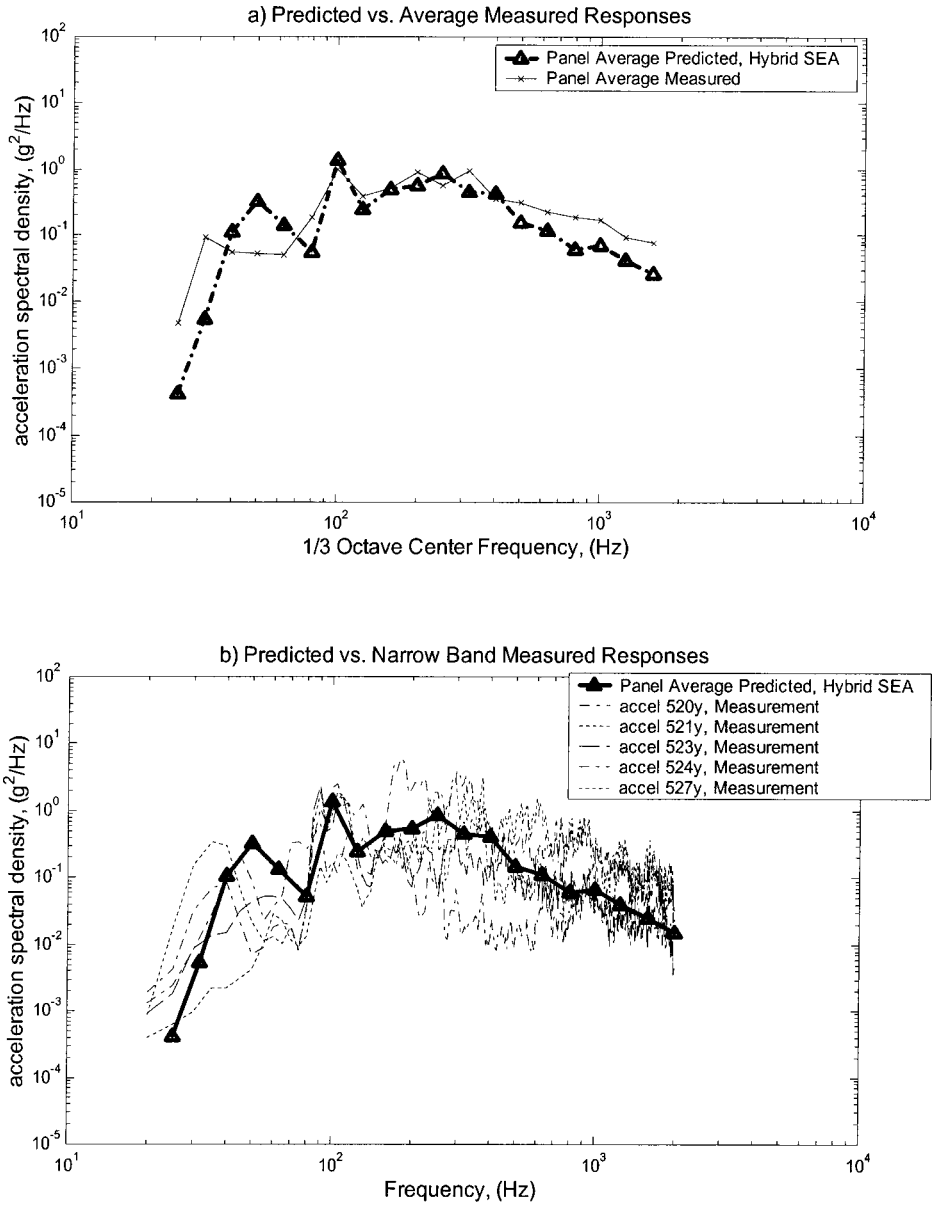


FIG. 12. Lumped mass loaded panel SEA hybrid prediction: (a) average predicted versus average measured, (b) average predicted versus individual measurements.

results appear to approximately follow a simple linear (on the log–log scale) relationship. A regression analysis is performed on the pertinent portions (up to radiation efficiency values of approximately 1.0) of the data from Table IV configurations 2 and 4. A simple equation (for the baffled complex panel, loaded and unloaded) representing the curve fit results for the radiation efficiency is

$$\sigma_{\mathrm{rad}} = 0.47\left(\frac{k_a}{k_p}\right)^{2.24}, \quad \text{for } k_a < 1.5k_p,$$
$$\sigma_{\mathrm{rad}} = 1.0, \quad \text{for } k_a \geqslant 1.5k_p, \tag{22}$$

where k_a=acoustic wave number, and k_p=(unloaded) panel wave number.

The simple fit given in Eq. (22) will be used in subsequent SEA analyses. The Eq. (22) results are compared to the computed modal values in Fig. 11. Note this result is for a baffled panel and approximates the equipment panel use as part of a ''closed box'' satellite structure. Additional results have also been reported,[27] evaluating the panel radiation efficiency in both baffled and unbaffled conditions. These results showed below the critical frequency the baffled complex panel radiation efficiency (both the total panel structure and lumped mass-loaded panels) increases approximately proportional to the frequency where the unbaffled complex panels increase approximately proportional to the frequency squared. Also of note are the results that showed the removal of the baffle for the complex panels results in much less of a radiation efficiency reduction than for the uniform panel (more than 6 dB at 30 Hz).

D. Summary of FE/BE results

The results of the finite element analysis show the nonhomogeneous features of this type of complex loaded panel have very significant effects on the built-up panel vibration characteristics. These effects are not readily approximated at all frequencies of concern with any of the SEA asymptotic approaches. However, at high frequencies (above approximately 1000 Hz) for all the configurations evaluated, the

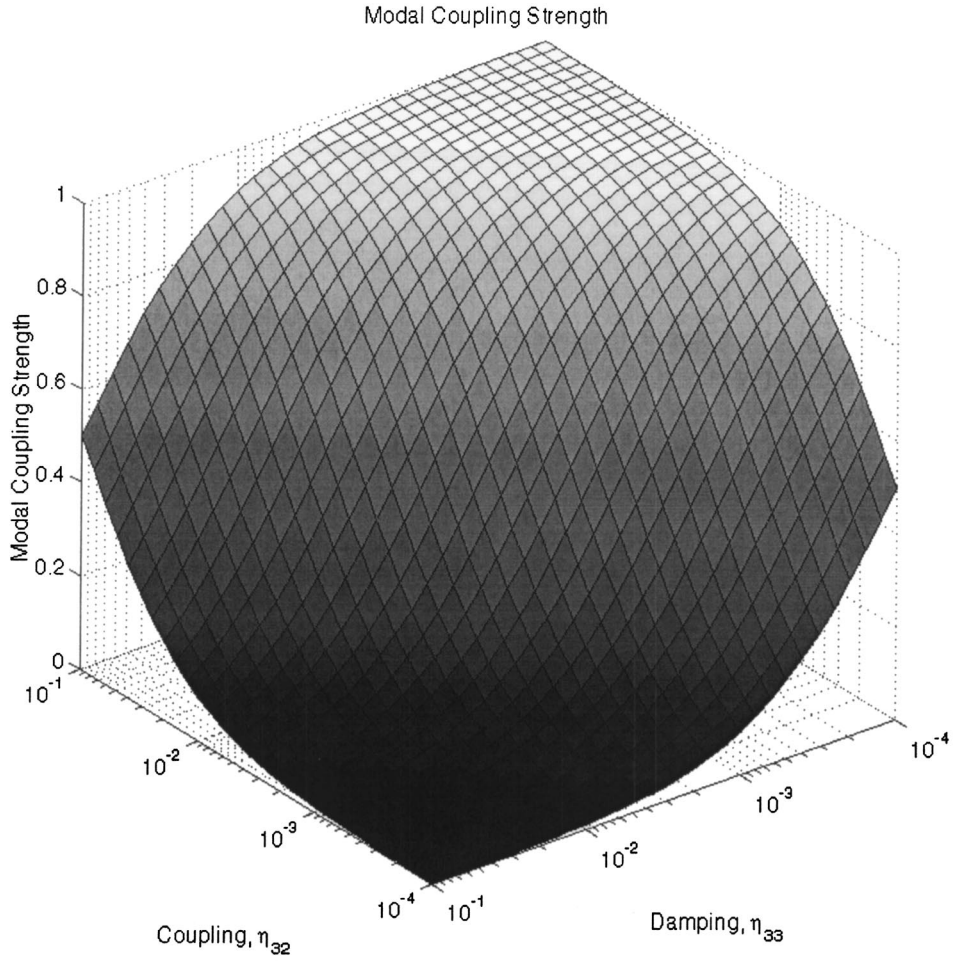


FIG. 13. Modal coupling strength versus coupling and damping loss factors.

FEA computed one-third octave modal density approaches the SEA value, which considers the stiffness of the panel face sheets and core loaded with the balance of the mass as limp (nonstructural). The results also show good agreement, for both the uniform as well as nonhomogeneous panels, between the actual mode count derived modal densities and the modal densities derived via the spatial average conductance.

The results from the BE analysis have characterized the baffled radiation efficiencies for the uniform and complex panel designs. The complex panel results show an approximate linear on log–log slope up to approximately 1000 Hz and a shift-up of the apparent critical frequency. The radiation efficiency peaks at about the same frequency for both the complex panels, even though the lumped mass-loaded panel has more than twice the mass of the total panel structure. Below the critical frequency the baffled complex panel radiation efficiency (both the total panel structure and lumped mass-loaded panels) increases approximately proportional to the frequency.

The results presented clearly show that for complex equipment-loaded panels traditional SEA methods must be augmented with other methods such as FEA/BEA to properly characterize the effects of nonhomogeneous features on vibration and radiation characteristics.

E. Panel with lumped masses

The lumped mass-loaded panel SEA model and response equations are developed in the same fashion as was done for the equipment loaded panel, except now there are no resonant attachments and all the attached mass is additional mass associated with the built up panel, as listed in Table I. It is assumed in this case that the stiffness/mounting preload tension of the lumped mass mounting fasteners are adequate to prevent resonances associated with the masses from occurring below 2000 Hz. Following analogous steps as for the equipment-loaded panel, the lumped mass loaded panel response is

$$\langle v_2^2 \rangle = \left(\frac{\eta_{21}}{\eta_{22} + \eta_{21}} \right) \left(\frac{n_2}{n_1} \right) \left(\frac{V_1}{\rho_0 C_0^2} \right) \left(\frac{\langle p_1^2 \rangle}{M_2} \right), \tag{23}$$

where n_1 is calculated in Eq. (2): V_1, ρ_0, C_0 are from Table III, $\langle p_1^2 \rangle$ are one-third octave experimental values shown in Table II, η_{22} is the panel damping loss factor that is assumed to be 0.025 up to 250 Hz, then $0.025(250/f)^{1/2}$ above 250 Hz, $\eta_{21} = (\rho_0 C_0 A_2 / \omega M_2)\sigma_{\text{rad}}$ is the panel radiation loss factor, A_2 is the panel area, M_2 is now the total (lumped mass-) loaded panel mass in Table I, σ_{rad} is now the new equipment panel radiation efficiency approximation developed via FEA/BEA, and n_2 is now the total (lumped mass-) loaded panel modal density developed via the FEA computed $\langle G \rangle$.

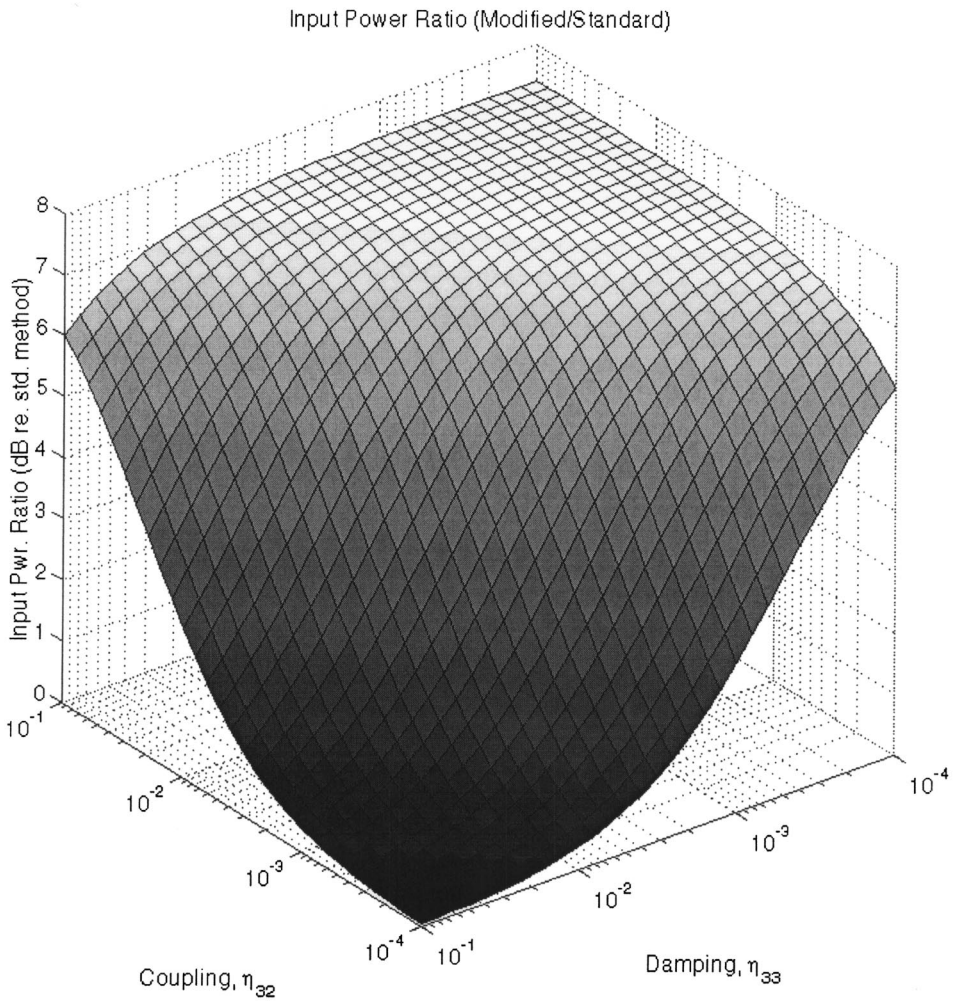


FIG. 14. Input power ratio modified/standard versus coupling and damping loss factors.

Figure 12 shows the SEA hybrid prediction along with the measured data for comparison. Overall, the results compare well. At higher frequencies the predicted average is somewhat lower than the measured data. This discrepancy is possibly due to biasing in the measured average (high) due to the limited acceleration measurements made and their positioning with respect to the attached masses. Many of the accelerometers may have been positioned locally away from the masses, thus at high frequencies sensing the motion of a ''locally unloaded'' panel. Another contributor to the high-frequency observation could be that the actual damping loss factor is significantly lower at high frequencies than the one utilized in the analysis. More data is required to fully resolve this discrepancy at high frequencies.

F. Panel with complex equipment

The results from the parameter investigation are now applied to the SEA model developed for the panel with multiple attached oscillators. As was done for the lumped mass-loaded panel, the modifications to the SEA model consist of incorporating the FEA/BEA results for the built-up panel modal density/conductance and radiation efficiency. In addition to these modifications, the external input power is also modified to account for the system internal couplings (panel to oscillators). Recent work[2] has shown the importance of addressing changes in the external input power due to internal couplings when dealing with equipment-loaded lightweight panels. The modified to standard input power ratio is

$$\frac{\pi_e}{\pi_e^0}=\left(1+\frac{n_3}{n_2^0}\xi_2^3\right)\left(1+\frac{M_3}{M_2^0}\xi_2^3\right)^{-1}, \tag{24}$$

where n_2^0 and M_2^0 are the modal density and mass of the master structure in isolation, $n=n_2^0+n_3\xi_2^3$ is the effective modal density[12] of the master structure *in situ* due to coupling to the adjunct structure, $M=M_2^0+M_3\xi_2^3$ is the effective mass[12] of the master structure *in situ* due to the coupling to the adjunct structure, and $\xi_2^3=E_3/n_3/E_2/n_2=\epsilon_3/\epsilon_2=\eta_{32}/(\eta_{33}+\eta_{32})$ is the ratio of modal powers (or modal energy ratios) defined as the modal coupling strength.

The modal coupling strength is plotted in Fig. 13 for a typical range of η_{32} and η_{33}. The modal coupling strength approaches unity as the coupling becomes very large relative to the damping in the adjunct structure and approaches zero as the coupling becomes very small relative to the adjunct structure damping. Maidanik[12] refers to Eq. (24) as ''the bridge between weak and strong coupling.'' For similar coupled structures, where $n_3/n_2\approx M_3/M_2$, the input power ratio is independent of the modal coupling strength ξ_2^3 and equal to unity (this is the assumption made in traditional

SEA). For a modally rich coupling, where $n_3/n_2 > M_3/M_2$, the input power ratio can be substantially more than unity when the modal coupling strength approaches one. Conversely when the coupling is mass rich, where $n_3/n_2 < M_3/M_2$, the input power ratio can take values substantially less than unity. The modified to standard input power ratio is plotted in Fig. 14 for modally rich coupling. These results illustrate the need to address internal coupling when estimating the external input power to strongly coupled dissimilar structures.

Applying the formulations developed to address the system internal couplings, the modified to standard input power ratio is as given in Eq. (24). Combining this with the expression for the standard input power[9] (plate power input from a diffuse acoustic field) gives the modified input power,

$$\pi_{12}^{\text{mod}} = \left(\frac{2\pi^2 C_0^2 \sigma_{\text{rad}} \langle p_1^2 \rangle n_2^0}{\omega^2 (M_2^0/A_2)}\right)\left(1+\frac{n_3}{n_2^0}\xi\right)\left(1+\frac{M_3}{M_2^0}\xi\right)^{-1}. \tag{25}$$

The complex equipment panel response ratio (modified/standard) can be written as

$$\frac{E_2^{\text{mod}}}{E_2^{\text{std}}} = \frac{\left(\frac{\eta_{21}^{\text{mod}}}{\eta_t+\eta_{21}^{\text{mod}}}\right)\left(\frac{n_2}{n_1}\right)E_1}{\left(\frac{\eta_{21}}{\eta_t+\eta_{21}}\right)\left(\frac{n_2^0}{n_1}\right)E_1}. \tag{26}$$

Equation (26) reduces to

$$\frac{E_2^{\text{mod}}}{E_2^{\text{std}}} = \frac{n_2}{n_2^0}\frac{(M_2^0\eta_t+\psi)}{(M_2\eta_t+\psi)}, \tag{27}$$

where n_2 and M_2 are the effective panel modal density and mass from Eq. (24). The superscript ^{0} indicates the panel in isolation values, where the modal density is calculated via the FEA computed $\langle G \rangle$ (configuration 2 from Table IV). The apparent panel damping η_t is as defined in Eq. (15) (also using the FEA computed $\langle G \rangle$) and ψ is a convenience parameter defined as

$$\psi = \frac{\rho_0 C_0 A_2 \sigma_{\text{rad}}}{\omega}. \tag{28}$$

Note that the modified and standard formulations are assumed to have the same band-averaged panel radiation efficiency σ_{rad}. For relatively small and large values of ψ, the panel response ratio in Eq. (27) reduces to

$$\frac{E_2^{\text{mod}}}{E_2^{\text{std}}} \approx \frac{n_2}{n_2^0}\frac{M_2^0}{M_2} = \frac{\pi_{12}^{\text{mod}}}{\pi_{12}^{\text{std}}}, \quad \text{for } \psi \ll M_2^0\eta_t,\ M_2\eta_t$$

and

$$\frac{E_2^{\text{mod}}}{E_2^{\text{std}}} \approx \frac{n_2}{n_2^0}, \quad \text{for } \psi \gg M_2^0\eta_t,\ M_2\eta_t.$$

Computation results show, for the parameters utilized in this analysis, that the first limit holds for the entire frequency range of interest. The modified panel response is now calculated as

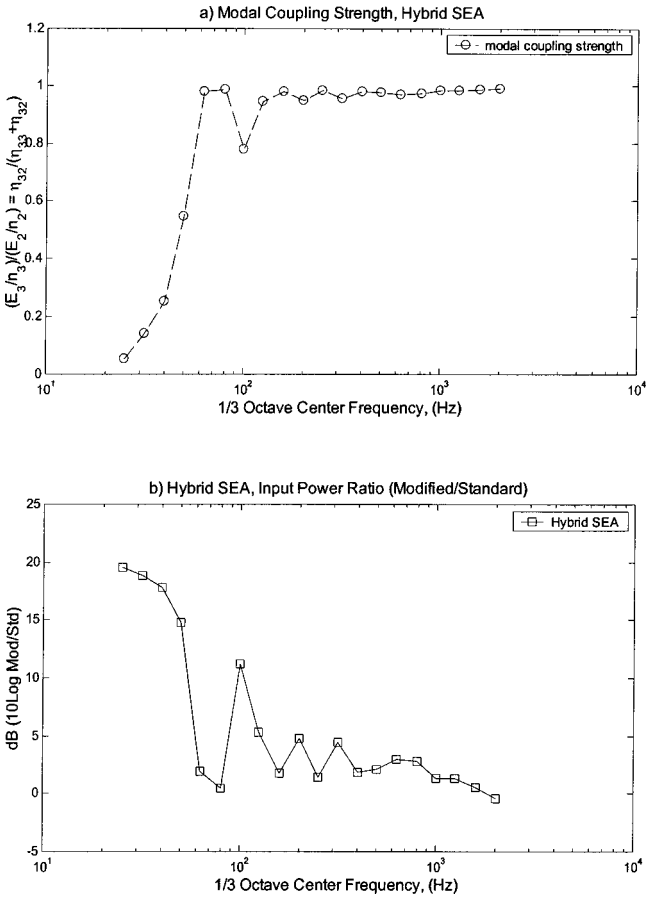


FIG. 15. SEA hybrid predictions: (a) modal coupling strength and (b) input power ratio.

$$\langle v_2^2 \rangle^{\text{mod}} = \frac{E_2^{\text{mod}}}{M_2^0}. \tag{29}$$

The loaded equipment panel-induced loss factor is as given in Eq. (15). The effective system loss factor is given in Eq. (10), where the modified power input is used along with the modified panel and oscillator energies.

The first condition run with the SEA hybrid (panel with coupled oscillators) model utilizes 100% of the equipment mass as resonant oscillator mass. Although unrealistic, this represents an upper limit and is evaluated first. Figures 15 and 16 show the modal coupling strength, input power ratio, and loss factors calculated for this configuration. Figure 15 shows that accounting for the internal couplings does significantly modify the input power over much of the frequency span of interest, particularly at the mid- to lower-frequencies, where the modal density ratio n_3/n_2 is highest. Figure 17 shows the individual portions of the SEA hybrid response predictions. The low-frequency (up to 125 Hz) prediction is based on applying the results from configuration 3 from Table IV (no equipment resonant) and the broadband results for the panel with oscillators come from configuration 2 from Table IV (all equipment resonant). Figure 18 combines the two panel predictions for a composite prediction and also shows the average oscillator response. The panel response is higher than the measured data above 200 Hz.

Since all the equipment mass is certainly not resonant

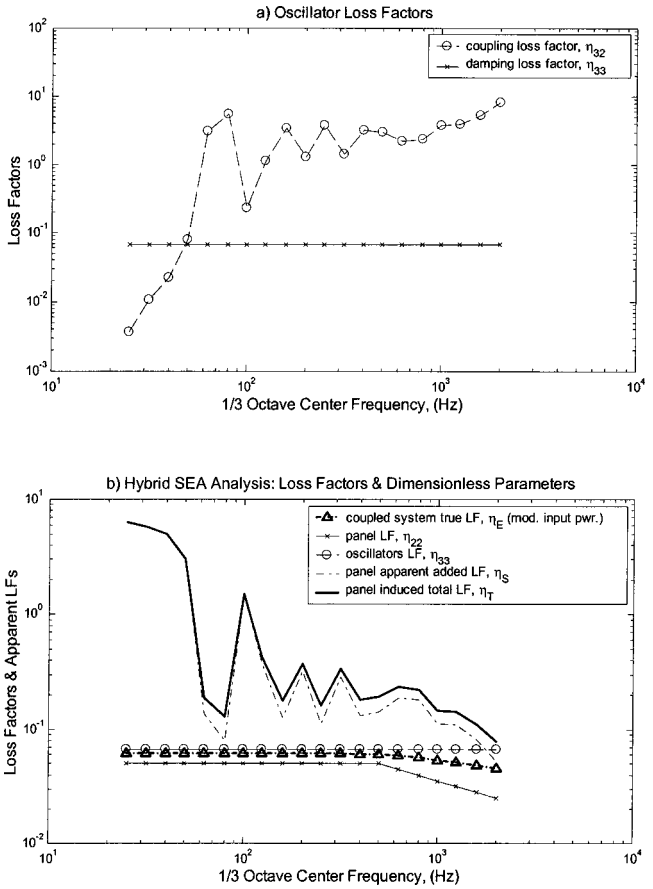


FIG. 16. SEA hybrid predictions loss factors: (a) oscillator coupling and damping loss factors, (b) a comparison of system, panel, oscillator, panel apparent, and panel-induced loss factors.

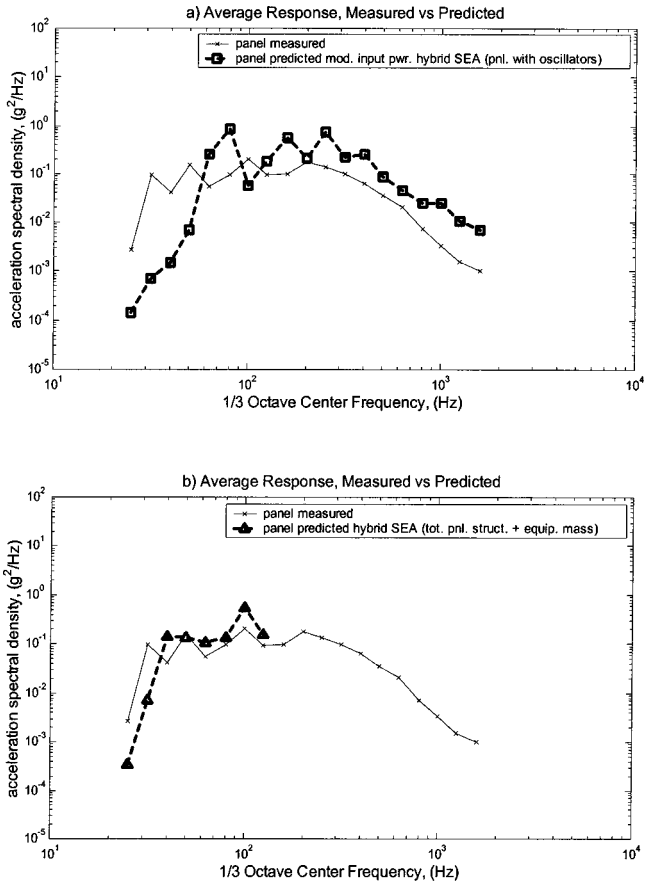


FIG. 17. SEA hybrid response predictions: (a) high frequency; (b) low frequency.

for the frequencies of interest, a scenario was evaluated where 25% of the equipment mass was assumed to be resonant. The final result for this case is shown in Fig. 19. The SEA hybrid result in Fig. 19 shows an improved correlation to the measured data, indicating the importance of the ratio of resonant to nonresonant attachment mass in the panel response. It should be noted that the portion of equipment mass assumed not to be resonant is now included with the effective panel mass when calculating the oscillator to panel coupling and panel response. The FEA computed total panel structure (case 2 from Table IV) modal density is utilized. The effects of the residual (nonresonant) part of the components on this modal density is an area for further investigation. As with the lumped mass-loaded panel, there is some discrepancy at the high frequencies. For the equipment-loaded panel the limited measurements were all made directly under critical electronic components; thus at high frequencies the residual mass of the component on the panel could significantly reduce the measured acceleration relative to a true spatial average. Therefore, as with the lumped mass-loaded panel, more measured data could aid in sorting out this apparent high-frequency discrepancy.

The SEA hybrid method predicted responses show good comparison to the measured data trends for both the lumped mass-loaded panel and the complex equipment-loaded panel. Figure 20 summarizes the SEA hybrid predictions and compares them to the measured data. The significant response trends between the two panels match well with the measured data. This has been accomplished by addressing explicitly the panel vibration, radiation, and attachment coupling in the analysis. Further experimentation will help to better characterize the measured mean levels and evaluate other critical parameters required to improve the models/predictions.

IV. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

The current design trends for communications satellites are toward higher power and physically larger structures. This results in more payload in the form of complex electronics mounted on large lightweight panels. These larger panels become more susceptible to acoustically induced high vibration responses during launch. Performing vibroacoustic analyses on these systems to evaluate and ensure design integrity during development is a necessity and the need to include the equipment explicitly in the analysis becomes more important. Typical industry standard methods utilize broad empirical generalizations for the equipment panel properties for incorporation into the SEA analysis and thus cannot discriminate between various types of attachment loading. Measurements for panels loaded with lumped masses (that have no modes) attached have been shown to be drastically different than those for the same panel loaded with complex equipment (that have many modes) attached. In this study we explicitly address the attachments and char-

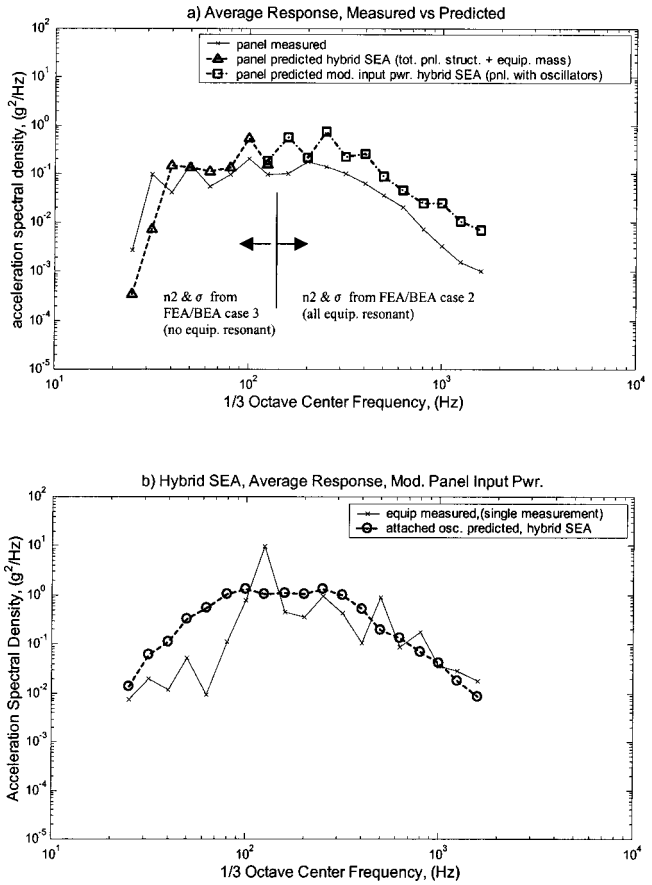


FIG. 18. SEA hybrid predictions versus measurements: (a) panel composite prediction, case 2 from Table IV with all equipment resonant at high frequencies; (b) attachment prediction.

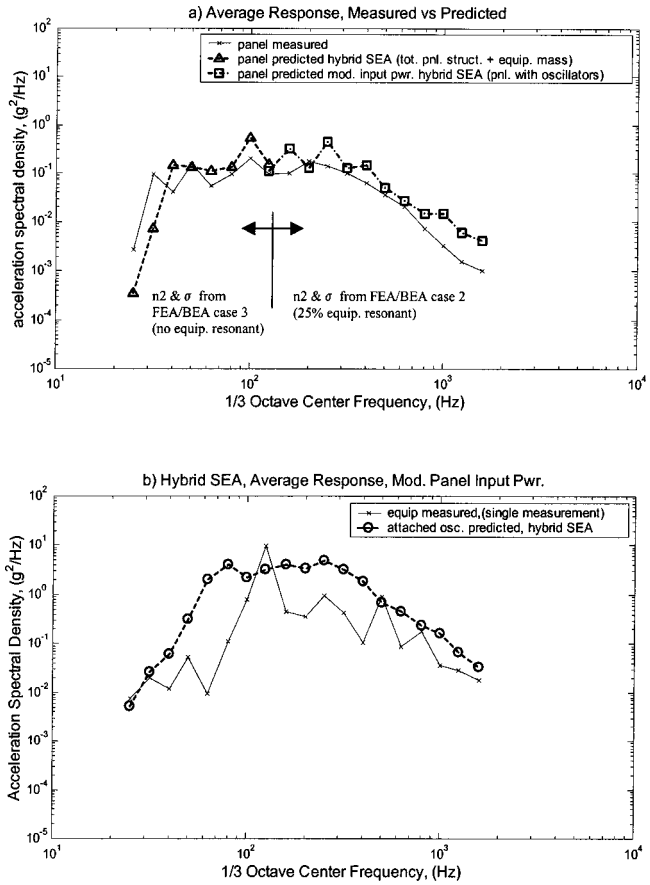


FIG. 19. SEA hybrid final composite prediction: (a) panel composite prediction, case 2 from Table IV with 25% equipment resonant at high frequencies, and (b) attachment prediction.

acterize how their properties affect the panel responses. An improved understanding of the vibration response mechanisms of lightweight complex equipment-loaded panels has been developed through the results of this study.

A SEA model for panels with multiple attachments was developed establishing the framework for further SEA hybrid model improvements. This SEA model also illustrated the significant effects of coupled or attached subsystems in terms of induced panel system damping. FEA and BEA methods were used to derive procedures for determining realistic parameters to input to the SEA hybrid models. Specifically, conductance/modal density and radiation efficiency for nonhomogeneous panel structures with and without mass loading were computed. The validity of using the spatially averaged conductance of panels with irregular features for deriving the structure modal density was also demonstrated. Maidanik's method of modifying the traditional SEA input power was implemented that illustrates the importance of accounting for system internal couplings when calculating the external input power. The modified input power/sensitivity analysis results showed the response prediction to be most sensitive to the dynamic mass ratio $M_{\text{equipment}}/M_{\text{panel}}$ and the attachment internal loss factor. The final response prediction results using the SEA hybrid models developed here show good agreement with the measured data trends.

In this study we have identified several key areas requiring further investigation. More rigorous measurements, from controlled experiments, of the loaded panel vibrations are required to evaluate/reduce any potential spatial bias errors in the averaged data. This is of special importance at the higher frequencies (above approximately 500 Hz for the panel considered here). This should also be done for panels with different attachment loadings (equipment spatial distribution and total mass) for analysis correlation and the establishment of empirical loading/response trends. Further experimental and analytical investigation is required to characterize the very critical dynamic mass ratio $M_{\text{equipment}}/M_{\text{panel}}$ as well as the modal density ratio $n_{\text{equipment}}/n_{\text{panel}}$. The initial decay rate methods used by Lalor[28] and Wu and Agren[29] may be useful for experimentally determining the equivalent or effective masses. Direct experimental determination of the in isolation damping loss factors for the panel and attachments as well as the *in situ* induced damping of the panel when coupled to the attachments would help to further validate the predictive approach derived here and are measurements that could be readily performed using the power injection method, as demonstrated by Bies and Hamid.[30] These loss factors would provide not only the separate subsystem loss characteristics but also the induced losses could be used to derive/infer the coupling loss characteristics of the attachments to the panel. Additional investigation/validation should be performed on the modified input power theory implemented in the analysis. A deterministic analysis or experiment utilizing much simpler attach-

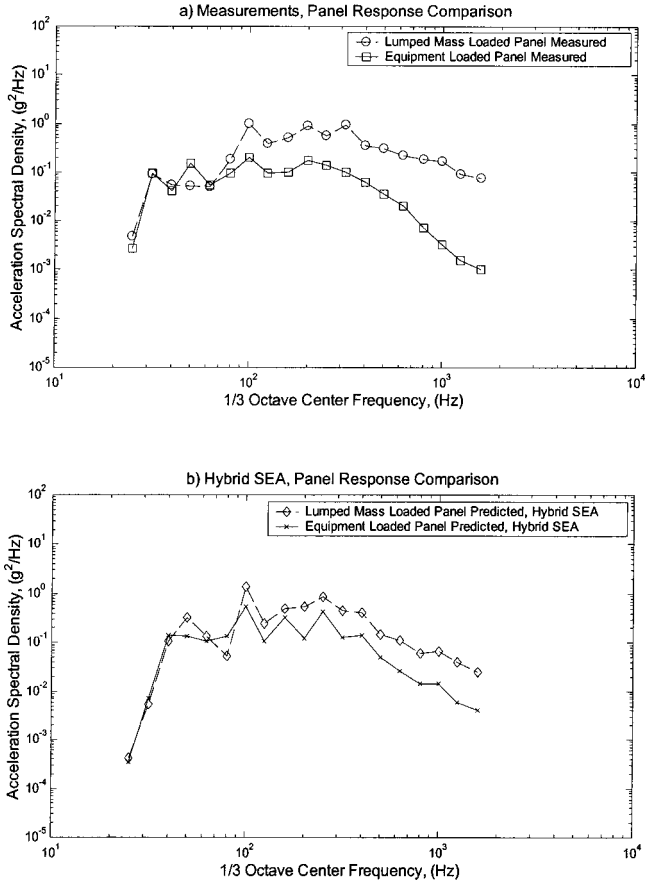


FIG. 20. Measured data versus predictions, lumped mass-loaded panel and equipment-loaded panel: (a) measured average responses, and (b) hybrid SEA predictions.

ments could help to validate the modified input power implementation. Drive point mobility measurements made on uniform and complex built-up panels would be useful for verifying the modal density results developed in this work. Not only are the built-up structure nonhomogeneous features of concern, the effects of which have been reported in the literature inconsistently, but, in addition, the effects of the honeycomb core cell structure reducing the anticipated measured modal density trends at high frequencies as touched upon by Clarkson and Ranky[31] and more recently by Renji.[32]

Additional investigation should be performed with respect to the oscillator model used for representing the resonant equipment. This model is of the simplest form and a more detailed evaluation may yield a more accurate way to describe the resonant attachments. One potential method for addressing the attachments may be as an empirically based ''fuzzy'' impedance term to add to the built-up panel system. Future planned experimentation will show whether this type of approach would be feasible.

Finally, one topic that has not been addressed in this work but nonetheless is of great interest and warrants significant future work is predicting the variability or confidence limits about the SEA predicted average responses. A study on the hardware evaluated in this investigation is currently underway[33] and is focusing on two important aspects of the response variability. Experimentally the ensemble complex panel response variability is being investigated as well as a computational evaluation of the band-averaged panel input power variability.

[1] C. Soize, ''Probabilistic structural modeling in linear dynamic analysis of complex mechanical systems: I. theoretical elements,'' Rech. Aerosp. **5**, 23–48 (1986).
[2] S. C. Conlon and S. A. Hambric, ''Statistical energy analysis prediction of the injected power and response of panels with multiple attachments,'' *Proceedings of Inter-Noise 99*, 6–8 December, Fort Lauderdale, FL, 1999.
[3] G. Maidanik, ''Induced damping by a nearly continuous distribution of nearly undamped oscillators: Linear analysis,'' J. Sound Vib. **240**, 717–731 (2001).
[4] S. C. Conlon, ''Response of non-homogeneous panels with complex attachments to diffuse acoustic loading,'' M.S. thesis, The Pennsylvania State University, August 2000.
[5] M. G. Smith, ''Validation of an SEA model of the Olympus satellite,'' European Space Agency Conference on Spacecraft Structures, Materials and Mechanical Testing, March 1996, Noordwijk, The Netherlands.
[6] J. N. Pinder and D. C. G. Eaton, ''Acoustically induced vibration of platform mounted equipment containing printed circuit boards,'' *Proceedings of International Conference on Spacecraft Structures and Mechanical Testing*, Noordwijk, The Netherlands, 24–26 April 1991.
[7] G. Maidanik, ''Response of ribbed panels to reverberant acoustic fields,'' J. Acoust. Soc. Am. **34**, 809–826 (1962); and corrections given in **57**, 1552 (1975).
[8] ''Structural acoustics design manual,'' European Space Agency, ESA PSS-03-204, March 1996.
[9] R. H. Lyon and R. G. DeJong, *Theory and Application of Statistical Energy Analysis*, 2nd ed. (Butterworth-Heinemann, Boston, 1995).
[10] J. H. Rindel, ''Dispersion and absorption of structure-borne sound in acoustically thick plates,'' Appl. Acoust. **41**, 97–111 (1994).
[11] G. Kurtze and B. G. Watters, ''New wall design for high transmission loss or high damping,'' J. Acoust. Soc. Am. **31**, 739–748 (1959).
[12] G. Maidanik and J. Dickey, ''On the external input power into coupled structures,'' *Proceedings of the IUTAM Symposium on Statistical Energy Analysis*, South Hampton, England, 1997.
[13] G. Maidanik, ''Power dissipation in a sprung mass attached to a master structure,'' J. Acoust. Soc. Am. **98**, 3527–3533 (1995).
[14] G. Maidanik and J. Dickey, ''Loss factors of pipe like structures containing beads,'' J. Acoust. Soc. Am. **99**, 2766–2774 (1996).
[15] G. Maidanik and J. Dickey, ''Design criteria for the damping effectiveness of structural fuzzies,'' J. Acoust. Soc. Am. **100**, 2029–2033 (1996).
[16] G. Maidanik and K. J. Becker, ''Various loss factors of a master harmonic oscillator coupled to a number of satellite harmonic oscillators,'' J. Acoust. Soc. Am. **103**, 3184–3195 (1998).
[17] M. E. McNelis, ''A modified VAPEPS method for predicting Vibroacoustic Response of Unreinforced Mass Loaded Honey Comb Panels,'' NASA-TM0101467, 1989.
[18] L. Lu, ''Dynamic Substructuring by FEA/SEA,'' in *Vehicle Noise*, Proceedings of the Winter Annual Meeting of the American Society of Mechanical Engineers, edited by S. H. Sung, K. H. Hsu, and R. F. Keltie (ASME, New York, 1990), NCA-Vol. 9, pp. 9–12.
[19] R. S. Langley and P. Bremner, ''A hybrid method for the vibration analysis of complex structural-acoustic systems,'' J. Acoust. Soc. Am. **105**, 1657–1671 (1999).
[20] P. Shorter, ''Combining finite elements and statistical energy analysis,'' Ph.D. thesis, University of Auckland, July 1998.
[21] J. E. Manning, ''Calculation of statistical energy analysis parameters using finite element and boundary element models,'' *International Congress on Recent Developments in Air- and Structure-Borne Sound and Vibration*, Auburn University, 6–8 March 1990.
[22] J. E. Manning, ''Use of measured mobility to improve SEA predictions in the mid-frequency range,'' *Proceedings of DETC99*, 17th ASME Biennial Conference on Mechanical Vibration and Noise, Las Vegas, Nevada, 1999.
[23] MSC/NASTRAN, Version 70.5.1 & Users Manual, The McNeal–Schwendler Corporation, Los Angeles, CA.
[24] L. Cremer, M. Heckl, and E. E. Ungar, *Structure-Borne Sound*, 2nd ed. (Springer-Verlag, Berlin, 1988).
[25] M. P. Norton, *Fundamentals of Noise and Vibration Analysis for Engineers* (Cambridge University Press, Cambridge, 1989).
[26] SYSNOISE Computational Vibro-Acoustics (Rev. 5.4 & Users Manual), LMS International, Leuven, Belgium.
[27] S. C. Conlon, S. A. Hambric, and J. E. Manning, ''Computational evalu-

ation of satellite equipment panel modal densities and radiation efficiencies,'' *Proceedings of NOISE-CON 2000*, joint ASA/NOISE–CON meeting, 3–8 December, Newport Beach, CA.
[28] N. Lalor, ''The experimental determination of vibrational energy balance in complex structures,'' *Proceedings of the SIRA Conference on Stress and Vibration*, London, 1989, Paper No. 108429.
[29] L. Wu and A. Angren, ''Analysis of initial decay rate in relation to estimates of loss factor and equivalent mass in experimental SEA,'' *Proceedings of IMSA 21*, Leuven, Belgium, 1996.
[30] D. A. Bies and S. Hamid, ''*In-situ* determination of loss and coupling loss factors by the power injection method,'' J. Sound Vib. **70**, 187–204 (1980).
[31] B. L. Clarkson and M. F. Ranky, ''Modal density of honeycomb plates,'' J. Sound Vib. **91**, 103–118 (1983).
[32] K. Renji, ''Experimental modal densities of honeycomb sandwich panels at high frequencies,'' J. Sound Vib. **237**, 67–79 (2000).
[33] S. C. Conlon, S. A. Hambric, and J. E. Manning, ''Order from disorder: A case study of the effects of structural inhomogeneity on structural-acoustic interaction,'' *Proceedings of Noise-CON 2001*, 29–31 October, Portland, Maine.

On the significance of reflection coefficients produced by active surfaces bounding one-dimensional sound fields

Timothy W. Leishman
Department of Physics and Astronomy, Brigham Young University, Eyring Science Center, Provo, Utah 84602

Jiri Tichy
Graduate Program in Acoustics, The Pennsylvania State University, Applied Science Building, University Park, Pennsylvania 16802

(Received 13 March 2001; revised 10 December 2002; accepted 13 December 2002)

Active boundary surfaces intended to control reverberation or other characteristics of enclosed sound fields have often been investigated using plane wave tubes. This paper presents an analysis of actively terminated semi-infinite and finite-length plane wave tubes to provide needed clarification of the effects of these surfaces. By considering relationships between complex pressure-amplitude reflection coefficients and acoustic energy quantities, the investigation reveals that increases in reflection coefficient moduli at terminations do not always produce corresponding increases in total energy or energy flux in adjacent fields. These relationships are shown to depend upon physical properties of the acoustic spaces, sources, and source positions. The investigation also demonstrates how the impact of reflection coefficients with moduli exceeding unity may be easily misinterpreted. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1550924]

PACS numbers: 43.50.Ki, 43.20.Tb, 43.20.Mv [MRS]

I. INTRODUCTION

The concept of utilizing active devices (e.g., electroacoustic transducers driven by electronic controllers) as acoustically responsive elements of enclosure boundaries has been investigated by many authors.[1–14] In contrast to elements of artificial reverberation systems, which attempt to recreate or synthesize desired acoustical signatures in rooms, these elements are intended to directly alter acoustical properties of enclosure boundaries or interact with primary sound sources to control specific sound field quantities.

Individual active boundary elements have frequently been evaluated as terminations to plane wave tubes.[3,5–9,11,14] Unfortunately, certain physical characteristics of these systems have been overlooked and capabilities of active boundary surfaces have consequently been misconstrued. In particular, complex pressure-amplitude reflection coefficients produced by active terminations and their relationships to acoustic energy quantities have been misunderstood, and implications of reflection coefficients with moduli exceeding unity ($|R|>1$) have largely been left to supposition. Because their characteristics are germane to extended enclosures with active boundary surfaces and fully three-dimensional fields, a concrete exploration of basic plane wave systems is important as a basis for further investigation.

One interpretation of $|R|>1$, from an exploration of active boundary elements, indicates that it connotes an increase in intensity in the adjacent field.[5] Another interpretation, from a study involving adaptive control of acoustic impedance, indicates that it connotes a supply of energy to the field.[8] (This notion agrees in part with the thinking of Morse and Ingard, who inauspiciously suggest that an active boundary surface ''feeds'' energy into an adjacent medium containing sources, independent of its reflection coefficient.[15]) A third interpretation, from a study of active surfaces intended for use in reverberant rooms, suggests that it connotes a super-reflection that increases reverberation in the field.[11] On the surface these interpretations may appear unobjectionable. However, further analysis reveals that they either inadequately or inaccurately characterize this attribute of active boundary surfaces—in both the plane wave tube arrangements used by their authors and in more general applications. They may have been based upon extrapolations from properties of passive materials, a hypothetical condition of constant incident pressure, or other preconceptions.

The problem of establishing effects of active boundary surfaces involves three interrelated facets. First, motion of an active boundary in the presence of impinging sound waves can produce a wide range of reflection coefficients, far greater than that of passive materials. Second, if another source exists in an adjacent finite space, vibration of an active boundary surface causes interactions with this source and consequently affects the field in a way that influences acoustic pressure incident upon itself—and its own reflection coefficient. Third, active boundary surfaces have acoustic impedances and reflection coefficients that are functions not only of their own vibrations, but vibrations of other sources within the adjacent space. Because these characteristics are not generally considered in the evaluation of common passive boundary materials, they are easily overlooked or misunderstood. Clearly, active boundary surfaces cannot be judged in the same context as passive boundary surfaces—particularly when their reflection coefficient moduli exceed unity.

To explore these matters further, this paper presents an analysis of two actively terminated plane wave tubes: a semi-infinite tube and a finite-length tube. Each incorporates a uniformly vibrating cross-sectional piston as an active

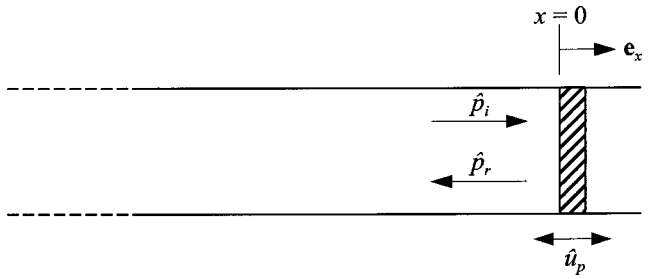


FIG. 1. Active vibrating piston terminating a semi-infinite plane wave tube. Constant incident pressure plane waves impinge upon the piston from a hypothetical source located at infinite distance to the left.

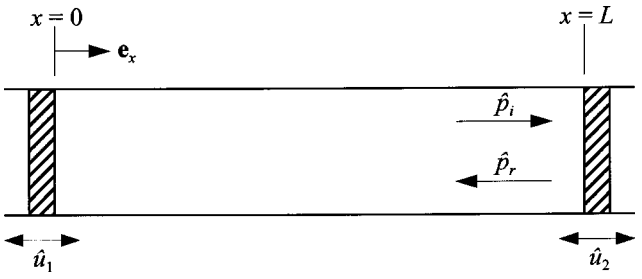


FIG. 2. Finite-length plane wave tube with active vibrating pistons at either end. The system behaves as an impedance tube driven by a primary source (piston 1) and terminated by an active boundary surface (piston 2).

boundary surface. Various sound field quantities are derived and relationships between their reflection coefficients and certain energy-related acoustic quantities are carefully examined. A few previously investigated aspects of the systems are briefly summarized in the analysis because of their pertinence.

II. ACTIVE PISTON TERMINATING A SEMI-INFINITE PLANE WAVE TUBE

Figure 1 depicts a lossless semi-infinite plane wave tube of cross-sectional area S terminated by an active uniformly vibrating piston. The piston vibrates freely but snugly with velocity $\mathbf{u}_p(t)=\mathrm{Re}\{\hat{u}_p \exp(j\omega t)\}\mathbf{e}_x$ inside the end of the tube.[16] This arrangement yields a convenient condition of constant incident pressure on the piston while precluding the need to address acoustic interaction with a primary source. The piston oscillates at the same angular frequency ($\omega = 2\pi f$) as the plane wave incident from the left, which is lower than the cutoff frequency of the first tube cross mode. The characteristic fluid impedance in the tube is $\rho_0 c$, where ρ_0 is the ambient fluid density and c is the speed of sound.

The specific acoustic impedance and pressure–amplitude reflection coefficient posed to the sound field by the vibrating piston may be represented as

$$Z_S=\frac{\hat{p}_B}{\hat{u}_p}-\rho_0 c \tag{1}$$

and

$$R=1-\rho_0 c\frac{\hat{u}_p}{\hat{p}_i}, \tag{2}$$

where $\hat{p}_B=2\hat{p}_i$ is the blocked surface pressure on the piston face. The piston velocity accordingly controls both the impedance and the degree of sound reflection, easily producing values of $|R|>1$.

Active acoustic intensity I in the tube may be expressed in terms of incident and reflected intensities or equivalently in terms of the incident intensity and the reflection coefficient:[17]

$$I=I_i-I_r=I_i(1-|R|^2). \tag{3}$$

Total time-averaged energy density w, with potential and kinetic energy density components, may likewise be expressed using the reflection coefficient:

$$w=w_p+w_k=\frac{|\hat{p}_i|^2}{2\rho_0 c^2}(1+|R|^2)=\frac{I_i}{c}(1+|R|^2). \tag{4}$$

The intensity, the related sound power, and the energy density are all independent of distance from the piston.

III. ACTIVE PISTON TERMINATING A FINITE-LENGTH PLANE WAVE TUBE

A lossless finite-length tube with uniformly vibrating pistons at either end extends the analysis of the active boundary surface to include its interaction with another source and a completely enclosed sound field. Although this model has been used elsewhere for analytical studies,[18–23] it is used here with a different emphasis. Figure 2 depicts the system, which involves a tube of length L and cross-sectional area S. The termination is controlled to respond in various ways to the primary source. If the velocities of the two pistons are known, several interesting aspects of the system may be considered, including basic acoustic quantities and sound field control.

A. Basic acoustic quantities

If pistons 1 (left) and 2 (right) vibrate with velocities $\mathbf{u}_1(t)=\mathrm{Re}\{\hat{u}_1 \exp(j\omega t)\}\mathbf{e}_x$ and $\mathbf{u}_2(t)=-\mathrm{Re}\{\hat{u}_2 \times \exp(j\omega t)\}\mathbf{e}_x$, respectively,[24] the boundary conditions yield the following relationships for acoustic pressure and particle velocity:

$$\hat{p}(x)=-j\rho_0 c\hat{u}_1\left[\sin kx+\left(\cot kL+\frac{\hat{u}_2}{\hat{u}_1}\csc kL\right)\cos kx\right], \tag{5}$$

$$\hat{\mathbf{u}}(x)=\hat{u}_1\left[\cos kx-\left(\cot kL+\frac{\hat{u}_2}{\hat{u}_1}\csc kL\right)\sin kx\right]\mathbf{e}_x \tag{6}$$

or

$$\hat{p}(x)=\frac{-j\rho_0 c}{2\sin kL}[(\hat{u}_1 e^{jkL}+\hat{u}_2)e^{-jkx}+(\hat{u}_1 e^{-jkL}+\hat{u}_2)e^{jkx}], \tag{7}$$

$$\hat{\mathbf{u}}(x)=\frac{-j}{2\sin kL}[(\hat{u}_1 e^{jkL}+\hat{u}_2)e^{-jkx} -(\hat{u}_1 e^{-jkL}+\hat{u}_2)e^{jkx}]\mathbf{e}_x, \tag{8}$$

where $k=\omega/c$ is the acoustic wave number. These results may be used to show that the specific acoustic impedance posed to the field by the active termination (piston 2) depends not only upon its own velocity, but upon the velocity of the primary source (piston 1):

$$Z_{S,2}=j\rho_0 c\left(\cot kL+\frac{\hat{u}_1}{\hat{u}_2}\csc kL\right) \tag{9}$$

(compare Ref. 22).

The reflection coefficient at $x=L$ may be obtained in a straightforward manner from this impedance or by using Eq. (7) with the relationship $R=\hat{p}_r/\hat{p}_i$:

$$R|_{x=L}=\frac{\hat{u}_1+\hat{u}_2 e^{jkL}}{\hat{u}_1+\hat{u}_2 e^{-jkL}}. \tag{10}$$

Because R also depends upon both $\hat{u}_1$ and $\hat{u}_2$, one may modify its value to achieve a desired degree of reflection or absorption at the termination by adjusting *either* piston velocity. For example, if total absorption $(R=0+j0)$ is required, the velocity condition is simply $\hat{u}_2=-\hat{u}_1\times\exp(-jkL)$.[20,21] If R is constrained in this fashion, the condition eliminates steady-state pressure reflected from the termination. Steady-state pressure *incident* upon the termination is eliminated if the velocities satisfy the condition $\hat{u}_2=-\hat{u}_1\exp(jkL)$.

Equations (5)–(8) also lead to a relationship for intensity. The components incident upon and reflected from the active termination are

$$I_i=\frac{\rho_0 c}{8\sin^2 kL}|\hat{u}_1+\hat{u}_2 e^{-jkL}|^2, \tag{11}$$

$$I_r=\frac{\rho_0 c}{8\sin^2 kL}|\hat{u}_1+\hat{u}_2 e^{jkL}|^2 \tag{12}$$

which satisfy the relationship $I_r/I_i=|R|^2_{x=L}$. The total intensity is

$$\begin{aligned} I=I_i-I_r&=\frac{\rho_0 c}{2\sin kL}|\hat{u}_1|^2\ \mathrm{Im}\left\{\frac{\hat{u}_2}{\hat{u}_1}\right\}\\ &=\frac{\rho_0 c}{2\sin kL}(u_{1,I}u_{2,R}-u_{1,R}u_{2,I}), \end{aligned} \tag{13}$$

where the subscripts R and I denote real and imaginary components, respectively. Notably, each of these intensity expressions depends upon both piston velocities and is independent of position x. However, total intensity is nonzero only if *both* piston velocities are nonzero and the velocity ratio $\hat{u}_2/\hat{u}_1$ has an imaginary component. The corresponding sound power quantities, related by the tube cross-sectional area S, behave similarly.

Energy density in the tube is given by the expression

$$\begin{aligned} w=w_p+w_k&=\frac{|\hat{p}(x)|^2}{4\rho_0 c^2}+\frac{\rho_0|\hat{u}(x)|^2}{4}\\ &=\frac{\rho_0}{4\sin^2 kL}[|\hat{u}_1|^2+|\hat{u}_2|^2+2(u_{1,R}u_{2,R}\\ &\quad+u_{1,I}u_{2,I})\cos kL]. \end{aligned} \tag{14}$$

Because of its spatial uniformity, total acoustic energy in the tube is simply

$$\begin{aligned} E=E_p+E_k=\frac{\rho_0 SL}{4\sin^2 kL}&[|\hat{u}_1|^2+|\hat{u}_2|^2\\ &+2(u_{1,R}u_{2,R}+u_{1,I}u_{2,I})\cos kL]. \end{aligned} \tag{15}$$

Another quantity of special interest is the potential energy component of total energy

$$\begin{aligned} E_p&=S\int_0^L\frac{|\hat{p}(x)|^2}{4\rho_0 c^2}dx\\ &=\frac{\rho_0 SL}{8\sin^2 kL}\Bigg[(|\hat{u}_1|^2+|\hat{u}_2|^2)\\ &\quad\times\left(\cos kL\frac{\sin kL}{kL}+1\right)+2(u_{1,R}u_{2,R}+u_{1,I}u_{2,I})\\ &\quad\times\left(\cos kL+\frac{\sin kL}{kL}\right)\Bigg]. \end{aligned} \tag{16}$$

B. Sound field control

Because active boundary surfaces may be used to control specific sound field quantities,[21,25] a few pertinent examples are discussed here as a basis for later discussion of how these control schemes relate to termination reflection coefficients. To minimize total pressure at a position x in the tube, the required relationship between $\hat{u}_1$ and $\hat{u}_2$ may be determined by setting $\hat{p}(x)=0$ in Eq. (5):

$$\hat{u}_2=-\hat{u}_1(\tan kx\ \sin kL+\cos kL). \tag{17}$$

At very low frequencies, this relationship produces substantial pressure reduction throughout the entire tube, but at higher frequencies it can lead to pressure increases at other positions. A more useful objective might then be to achieve consistent *global* control of the field. A common approach to this problem for enclosed fields involves the minimization of total potential energy E_p.[21] Another advantageous approach is to minimize total energy density w at discrete positions.[26,27]

Given a fixed velocity $\hat{u}_1$, the total potential energy in the enclosure is minimized by the termination piston velocity

$$\hat{u}_2=-\hat{u}_1\left(\frac{\cos kL+\sin kL/kL}{\cos kL\sin kL/kL+1}\right) \tag{18}$$

(compare Ref. 20). The velocity required to minimize total energy density at any point within the enclosure is

$$\hat{u}_2 = -\hat{u}_1 \cos kL \tag{19}$$

which is the same as that required to control total energy E within the tube.[20,21] The solutions to control E_p and w converge when $kL \to n\pi$, where n is a non-negative integer, and more generally when $kL \gg \pi$. Several other energy control schemes could also be investigated using the model.[14]

IV. ENERGY-RELATED EFFECTS OF REFLECTION COEFFICIENTS PRODUCED BY ACTIVE TERMINATIONS

Having considered several fundamental characteristics of active terminations and adjacent plane wave tube sound fields, the discussion now turns to the central focus of the paper: the significance of reflection coefficients generated by these active surfaces. The preceding sections have demonstrated that reflection coefficient moduli of active boundary surfaces can easily exceed unity (i.e., $|R|>1$), the common upper bound for passive materials. On the other hand, they can easily replicate moduli of passive materials (i.e., $0 \leqslant |R| \leqslant 1$). This section interprets this wide range of reflection coefficients and its relationships to acoustic energy quantities in adjacent fields. It also reviews the past interpretations of $|R|>1$, demonstrating how their seemingly intuitive descriptions can be misleading.

A. Semi-infinite tube

When a piston terminating a semi-infinite tube (Fig. 1) is controlled to produce a reflection coefficient modulus of $|R|=0$, it produces no reflected intensity. From Eq. (3), it is clear that the total intensity becomes $I=I_i$ and from Eq. (4), the energy density becomes $w=I_i/c$. In contrast, when the piston is controlled to produce a reflection coefficient modulus of $|R|=1$, the reflected intensity magnitude equals the incident intensity magnitude ($I_r=I_i$), so from Eq. (3) the total intensity vanishes. From Eq. (4), the total energy density becomes double that present under the prior condition, i.e., $w=2I_i/c$. Thus, in comparison to an anechoic condition ($R=0+j0$), the imposition of a rigid termination ($R=1+j0$) increases energy in the field. Yet it simultaneously eliminates total sound-energy flux. This simple example demonstrates an important point: when one discusses boundary surfaces and their energy-related effects in an adjacent space, one must explicitly state both the specific energy quantity being considered and a comparative boundary condition. Because the two extreme passive conditions mentioned here provide useful benchmarks to compare active boundary surfaces, energy *ratios* that involve them are quite useful. Specifically, ratios of total intensity, reflected intensity, and total energy density (with arbitrary R) to the respective quantities determined under anechoic or rigid termination conditions merit consideration.

The useful intensity ratios for the semi-infinite tube are total intensity to that present under the anechoic condition and reflected intensity to that present under the rigid condition

$$\frac{I}{I_{an}} = 1 - |R|^2, \tag{20}$$

$$\frac{I_r}{I_{r,rg}} = |R|^2. \tag{21}$$

The ratios involving total energy density are

$$\frac{w}{w_{an}} = 1 + |R|^2, \tag{22}$$

$$\frac{w}{w_{rg}} = \frac{1}{2}\,(1 + |R|^2). \tag{23}$$

A number of conclusions may be drawn from these simple expressions and the related expressions of Sec. II. First, when $0<|R|<1$, total intensity remains positive. Sound-energy flux is directed toward the termination so that the piston absorbs at least a portion of the incident energy. Energy density in the field is greater than in the purely anechoic case, but less than in the purely rigid case. When $|R|>1$, the reflected intensity magnitude exceeds the fixed incident intensity magnitude and therefore the reflected intensity magnitude encountered under the rigid boundary condition. Total intensity becomes negative so sound-energy flux is directed away from the piston into the adjacent space. Both the vector magnitude of the total intensity and the energy density become greater than in either the anechoic or rigid case, growing in proportion to $|R|^2$. The active boundary with $|R|>1$ therefore *increases* both energy flow back to the field and total energy per unit tube length.

The past interpretations for $|R|>1$ mentioned in the introduction coincide well with this model and follow trends of passive materials. The first interpretation suggests that large reflection coefficient moduli create an increase in intensity. Its authors specifically indicate that $|R|>1.5$ provides "more than [a] doubling [of] acoustic intensity."[5] Significantly, this assertion appears plausible only if the indicated intensity refers to reflected intensity relative to that present under a rigid boundary condition [see Eq. (21)] or relative to incident intensity (from $I_r/I_i=|R|^2_{x=L}$). The second and third interpretations are also plausible. Nevertheless, the following section reveals deficiencies of the interpretations under more general conditions.

B. Finite-length plane wave tube

Energy-related quantities and ratios for the two-piston plane wave tube model (Fig. 2) may also be expressed in terms of the active termination reflection coefficient. However, a few general characteristics of the reflection coefficient modulus $|R|_{x=L}$ should first be considered.[28] As shown in Eq. (10), the complex reflection coefficient is a function of both the primary and secondary piston velocities. Its modulus may be expressed in terms of the complex velocity ratio $\hat{u}_2/\hat{u}_1$:

$$|R|_{x=L} = \left(\frac{1 + (\hat{u}_2^*/\hat{u}_1^*)e^{-jkL} + (\hat{u}_2/\hat{u}_1)e^{jkL} + |\hat{u}_2/\hat{u}_1|^2}{1 + (\hat{u}_2^*/\hat{u}_1^*)e^{jkL} + (\hat{u}_2/\hat{u}_1)e^{-jkL} + |\hat{u}_2/\hat{u}_1|^2} \right)^{1/2}. \tag{24}$$

From this relationship, it can be demonstrated that $\hat{u}_2/\hat{u}_1$ must have an imaginary component as a necessary condition for $|R|_{x=L} \neq 1$. This result is reminiscent of the condition for

nonzero active intensity discussed in Sec. III A. Interestingly, as demonstrated in Sec. III B, there are many control conditions that simply do not satisfy this requirement. For example, whether one controls acoustic pressure at a point, energy density, total energy, or total potential energy, the resulting reflection coefficient modulus remains the same: $|R|_{x=L}=1$.

Total intensity in the tube, expressed in terms of the termination reflection coefficient, is obtained from Eqs. (10) and (13):

$$I=\frac{\rho_0 c}{2}|\hat{u}_1|^2\left\{\frac{1-|R|^2_{x=L}}{|R|^2_{x=L}+1-2[(R_R|_{x=L})\cos 2kL+(R_I|_{x=L})\sin 2kL]}\right\}. \tag{25}$$

As with the semi-infinite tube, the useful intensity ratios are total intensity to that present under the anechoic condition and reflected intensity to that present under the rigid condition:

$$\frac{I}{I_{an}}=\frac{1-|R|^2_{x=L}}{|R|^2_{x=L}+1-2[(R_R|_{x=L})\cos 2kL+(R_I|_{x=L})\sin 2kL]}, \tag{26}$$

$$\frac{I_r}{I_{r,rg}}=\frac{4|R|^2_{x=L}\sin^2 kL}{|R|^2_{x=L}+1-2[(R_R|_{x=L})\cos 2kL+(R_I|_{x=L})\sin 2kL]}. \tag{27}$$

These expressions demonstrate that in the limit of large $R|_{x=L}$, i.e., $|R|_{x=L}\gg 1$, the ratio $I/I_{an}\to -1$ while the ratio $I_r/I_{r,rg}\to 4\sin^2 kL$. These results are important departures from the analogous results for the semi-infinite plane wave tube. Total intensity magnitude approaches that encountered under the *anechoic* boundary condition, but its sense is reversed (i.e., sound-energy flux is directed away from the termination and toward the enclosure and primary piston). Reflected intensity differs by a frequency-dependent factor $4\sin^2 kL$ when compared to that present under the rigid boundary condition.

Figure 3 shows a surface plot of I/I_{an} as a function of $R_R|_{x=L}$ and $R_I|_{x=L}$ for the frequency $f=9c/16L$. The rise between consecutive contour lines represents a ratio increment of 0.1. As the plot indicates, the ratio diverges both positively and negatively on either side of a single point on the complex $R|_{x=L}$ plane. At several other points, including those for which $|R|_{x=L}\gg 1$, it approaches a value of -1, as evidenced by the superposed mesh representing the $I/I_{an}=-1$ plane. (The divergences are truncated in the figure to focus attention on the behavior of the ratio near this plane.) Figure 4 provides additional insight through a closer overhead view of the contours. The heavy dashed circle indicates the values for which $|R|_{x=L}=1$. Notably, the eccentric contour rings become progressively smaller and closer together in the divergent regions with centers gradually approaching a

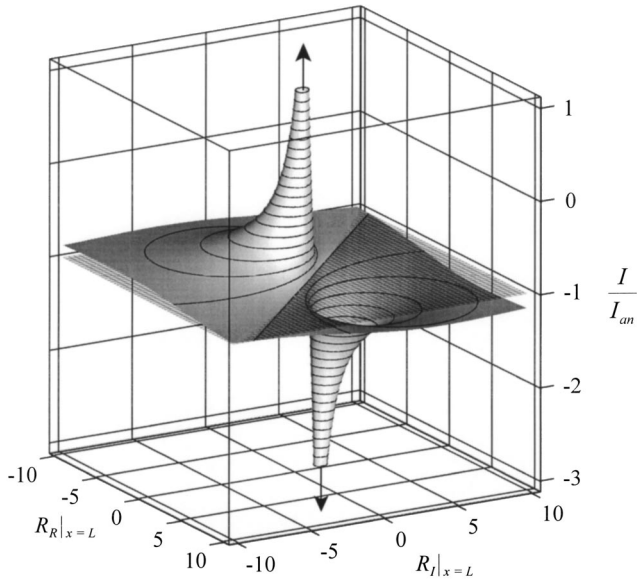


FIG. 3. Surface plot showing the ratio of total intensity in the two-piston plane wave tube to that present under an anechoic termination condition. The ratio is plotted as a function of the real and imaginary parts of the termination reflection coefficient, $R_R|_{x=L}$ and $R_I|_{x=L}$, for the frequency $f=9c/16L$. Contour lines are shown at increments of 0.1. The superposed mesh represents the $I/I_{an}=-1$ plane.

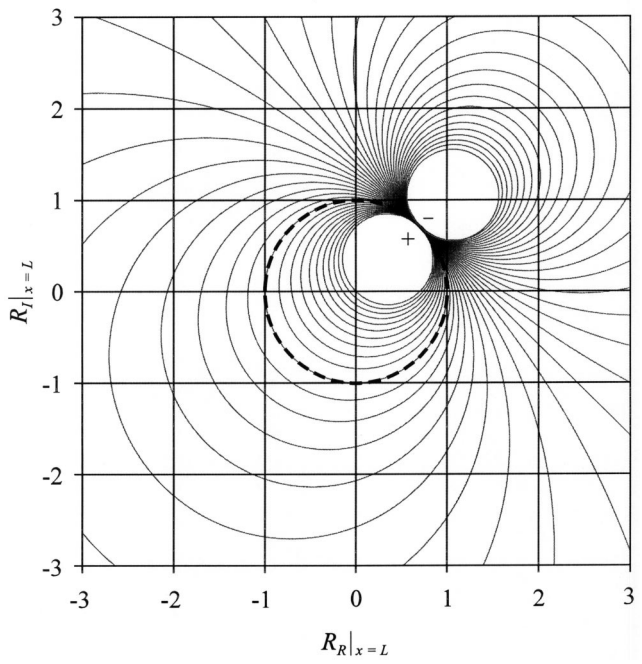


FIG. 4. Contour plot showing the ratio of total intensity in the two-piston plane wave tube to that present under an anechoic termination condition. The ratio is plotted as a function of the real and imaginary parts of the termination reflection coefficient, $R_R|_{x=L}$ and $R_I|_{x=L}$, for $f=9c/16L$. Contour lines are shown at increments of 0.1 with truncation at $I/I_{an}=1$ and -3 (see Fig. 3). The heavy dashed circle delineates $|R|_{x=L}=1$.

single point on this unit circle. Rings accumulating inside the circle correspond to the positive divergence, whereas those accumulating outside the circle correspond to the negative divergence. The series of rings are truncated at $I/I_{an}=1$ and $I/I_{an}=-3$, respectively, corresponding to the surface plot truncations of Fig. 3. Otherwise, they would become progressively smaller and closer together until they converged to the point $R|_{x=L}=\sqrt{2}/2+j\sqrt{2}/2$. This point and the divergence pair rotate counterclockwise around the unit circle with increasing frequency.

Further inspection of these figures and Eqs. (25) and (26) reveals that as long as $|R|_{x=L}<1$, any value of $R|_{x=L}$ yields a value of $I/I_{an}>0$; net sound-energy flux is directed toward the termination for all values of kL so that the vibrating piston consistently absorbs at least a portion of the steady-state incident intensity. Interestingly, if $R|_{x=L}$ is held constant in this range, the frequency-averaged total intensity (averaged over integer multiples of kL/π or over an interminably wide bandwidth) remains identical to that encountered under the anechoic boundary condition. In contrast, if $|R|_{x=L}>1$, total sound-energy flux is consistently directed away from the termination, toward the enclosed sound field and the primary piston. If $R|_{x=L}$ is held constant in this range, the frequency-averaged total intensity is equal in magnitude, but opposite in direction to that present under the anechoic boundary condition.

As $|R|_{x=L}\to 1$, the total intensity ratio in Eq. (26) converges to zero for most kL, but periodically diverges with dependence upon complex $R|_{x=L}$. When $R|_{x=L}$ is held constant with a modulus of unity, the frequency-averaged intensity ratio tends to zero. The behavior of the reflected intensity ratio in Eq. (27) also depends upon complex $R|_{x=L}$ as $|R|_{x=L}\to 1$. If the coefficient contains an imaginary component, the ratio diverges periodically over kL. However, if the coefficient is purely real, it coincides with a rigid surface which must produce a ratio of unity.

An examination of total acoustic energy within the enclosure reveals further differences between the two-piston and semi-infinite plane wave tube models. Equations (10) and (15) yield the following relationship for total energy:

$$E=\frac{\rho_0 SL}{2}|\hat{u}_1|^2\left\{\frac{|R|^2_{x=L}+1}{|R|^2_{x=L}+1-2[(R_R|_{x=L})\cos 2kL+(R_I|_{x=L})\sin 2kL]}\right\}. \tag{28}$$

The ratios of total energy with arbitrary $R|_{x=L}$ to the total energies present under anechoic and rigid boundary conditions are then

$$\frac{E}{E_{an}}=\frac{|R|^2_{x=L}+1}{|R|^2_{x=L}+1-2[(R_R|_{x=L})\cos 2kL+(R_I|_{x=L})\sin 2kL]}, \tag{29}$$

$$\frac{E}{E_{rg}}=\frac{2\sin^2 kL(|R|^2_{x=L}+1)}{|R|^2_{x=L}+1-2[(R_R|_{x=L})\cos 2kL+(R_I|_{x=L})\sin 2kL]}. \tag{30}$$

In the limit of large $R|_{x=L}$, i.e., for $|R|_{x=L}\gg 1$, the ratio in Eq. (29) tends to unity. Thus, the total acoustic energy in the enclosure approaches that present under the anechoic boundary condition; there is no significant increase. One example of such behavior is mentioned in Sec. III A, wherein the termination is controlled to eliminate steady-state pressure incident upon it. Under this condition, the reflection coefficient modulus at the termination tends to infinity for all values of kL, except as $kL\to n\pi$. The termination velocity is adjusted relative to the primary source velocity in a manner that effectively causes the primary source to behave as an anechoic absorber. The total energy is therefore equal to that of the complementary condition in which the termination acts as the anechoic absorber. In addition, it is clear from Eq. (29) that for $|R|_{x=L}\ll 1$, total energy in the enclosure consistently approaches that present under the anechoic condition. Hence, large relative changes (either increases or decreases) in total enclosed energy occur only when the termination reflection coefficient modulus roughly approaches unity. Some of the control examples discussed in Sec. III B are representative of this behavior for energy decreases.

Enclosed energy characteristics are further clarified in Fig. 5. A surface plot of the energy ratio E/E_{an} is shown for

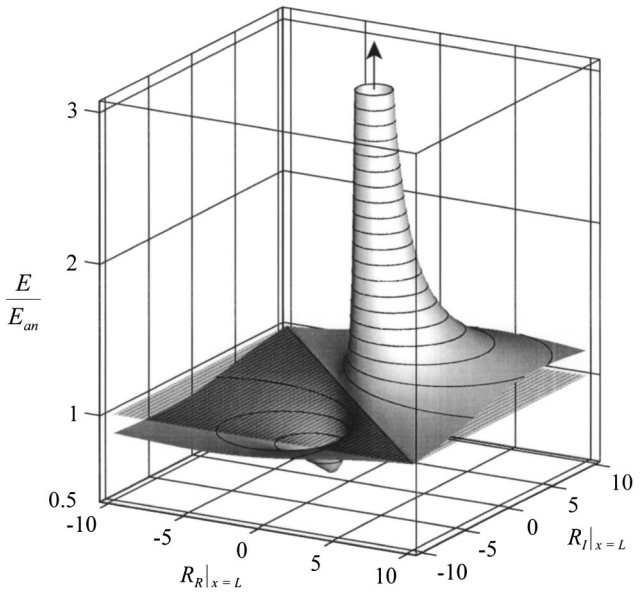


FIG. 5. Surface plot showing the ratio of total energy in the two-piston plane wave tube to that present under an anechoic termination condition. The ratio is plotted as a function of the real and imaginary parts of the termination reflection coefficient, $R_R|_{x=L}$ and $R_I|_{x=L}$, for $f=9c/16L$. Contour lines are shown at increments of 0.1. The superposed mesh represents the $E/E_{an}=1$ plane.

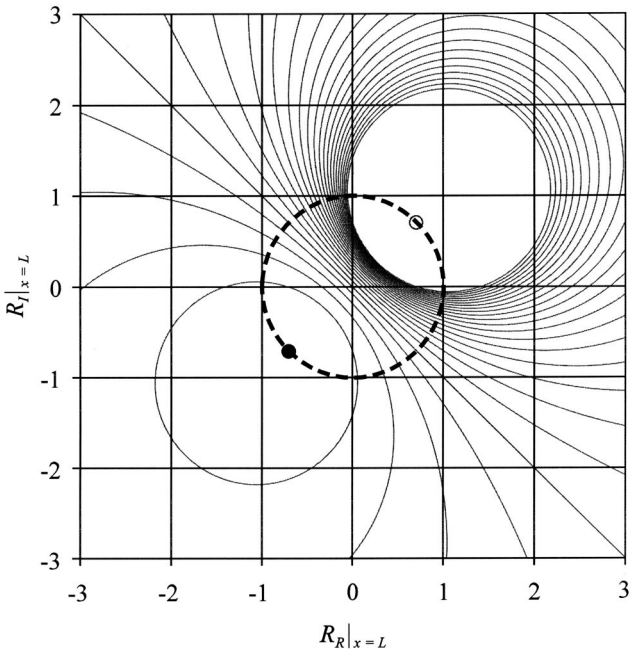


FIG. 6. Contour plot showing the ratio of total energy in the two-piston plane wave tube to that present under an anechoic termination condition. The ratio is plotted as a function of the real and imaginary parts of the termination reflection coefficient, $R_R|_{x=L}$ and $R_I|_{x=L}$, for $f=9c/16L$. Contour lines are shown at increments of 0.1 with truncation at $E/E_{an}=3$ (see Fig. 5). The heavy dashed circle delineates $|R|_{x=L}=1$. The small ring on the unit circle marks the point of maximum divergence, whereas the small black dot marks the point of the ratio minimum ($E/E_{an}=0.5$).

the frequency $f=9c/16L$ with a superposed mesh representing the $E/E_{an}=1$ plane. The rise between consecutive contour lines again represents a ratio increment of 0.1. The plot shows that the ratio diverges around one point on the complex $R|_{x=L}$ plane and converges to a minimum value of 0.5 at another. At several other points, including those for which $|R|_{x=L}\gg 1$, it approaches a value of 1. Figure 6 provides additional insight through a closer overhead view of the contours. The eccentric rings accumulating in the divergent region of the upper-right quadrant are truncated at $E/E_{an}=3$ to correspond to the surface plot truncation of Fig. 5. Otherwise, they would become progressively smaller and closer together until they converged to a single point encircled by the small ring on the unit circle: $R|_{x=L}=\sqrt{2}/2+j\sqrt{2}/2$. The eccentric rings accumulating less conspicuously in the lower-left quadrant correspond to the ratio minimum of 0.5. The inner-most contour ring shown for the region has a value of $E/E_{an}=0.6$. With finer contour increments, additional rings would become apparent that converge to the minimum point indicated by the black dot on the unit circle: $R|_{x=L}=-\sqrt{2}/2-j\sqrt{2}/2$. Accordingly, both the point of divergence and the ratio minimum fall on the $|R|_{x=L}=1$ circle, 180 degrees from each other. This pair rotates counterclockwise around the circle with increasing frequency.

It is important to recognize that under some conditions, reflection coefficients with moduli less than unity, i.e., $|R|_{x=L}<1$, can relatively increase total energy in the enclosed field. Purely real reflection coefficients in the range $0<R_R|_{x=L}<1$ increase energy relative to that encountered under the anechoic boundary condition but not relative to that encountered under the rigid boundary condition. However, purely imaginary or complex reflection coefficients, even with very small imaginary components in the range $0<R_I|_{x=L}<1$, can produce significant increases to total energy in comparison to that encountered under *either* benchmark condition. As these imaginary components become larger within this range, the increases may become particularly dramatic, even when averaged over frequency.

With these developments in mind, it is instructive to revisit the past interpretations for $|R|>1$ mentioned in the introduction—each coming from studies involving physical arrangements similar to the two-piston plane wave tube model.[29] It is clear from the present results that such large reflection coefficient moduli do *not* necessarily connote (1) an increase in an intensity quantity or (2) an addition of energy to the field in comparison to that encountered under anechoic or rigid benchmark conditions. While it is true that sound-energy flux is directed toward the adjacent field for $|R|>1$, energy is not supplied to the field in the sense that it necessarily *augments or maintains* the field energy relative to a benchmark condition. In essence, direction of flux is extraneous to the total enclosed energy. In some cases, values of $|R|>1$ relatively *diminish* total enclosed energy, whereas in other cases, values of $|R|<1$ relatively *increase* total enclosed energy. Furthermore, neither sound-energy flux nor total enclosed energy is generally proportional to $|R|$ or $|R|^2$. In fact, the analysis indicates that the most dramatic swings in these quantities occur as $|R|\to 1$. Thus, the concept of $|R|>1$ connoting a supply of energy to the adjacent field loses relevance. The interpretation that $|R|>1$ connotes a super-reflection that increases reverberation within the adjacent sound field does not generally hold if the term ''super-reflection'' describes reflected intensity magnitude exceeding that reflected from a rigid termination. However, if it describes reflected intensity magnitude exceeding incident intensity magnitude, the discussion in Sec. III A suggests that the concept does hold. Nevertheless, the total steady-state energy in the enclosure, which involves reflected energy, certainly does *not* always increase for values of $|R|>1$.

V. CONCLUSIONS

This work has demonstrated that reflection coefficients of active boundary surfaces and their effects on adjacent one-dimensional sound fields may be misinterpreted when important physical properties of the systems are overlooked. In the past, conclusions about these surfaces have coincided ostensibly with constant incident pressure models and behaviors of passive materials. This analysis has demonstrated several distinctions that broaden understanding of the surfaces and their potentially large reflection coefficients (i.e., $|R|>1$).

Investigation of a finite-length plane wave tube with a primary piston source and an active piston termination has produced several important results that clarify relationships between reflection coefficients and pertinent time-averaged energy quantities. First, for the active termination to create a value of $|R|$ greater or less than unity, or to allow a net flow of sound energy, its velocity must have an imaginary component relative to that of the primary source. Second, to understand the impact of reflection coefficients upon energy

flux or total enclosed energy in the tube, one must not only know the reflection coefficient modulus at the termination, but its real and imaginary components. Third, increases in reflection coefficient moduli at the termination *do not* necessarily indicate corresponding increases in sound-energy flux back toward the enclosed space or total enclosed energy. Fourth, energy-related effects of reflection coefficient moduli at the termination are frequency-dependent functions of the adjacent space and vibration of the primary source. Fifth, while reflection coefficient moduli exceeding unity may not produce *absorption* of energy at the termination, they can relatively *reduce* the vector magnitude of total intensity and total energy within the enclosed space. Sixth, very large reflection coefficients (i.e., $|R| \gg 1$), produce both a vector magnitude of total intensity and total enclosed energy approaching those present under an anechoic boundary condition (i.e., $|R| = 0$). Seventh, large relative changes (either increases or decreases) in total intensity and total enclosed energy occur only as active termination reflection coefficient moduli roughly approach unity.

From these findings one may conclude that sound-energy-related effects of active boundary surfaces are frequency-dependent functions of the spaces they bound and the acoustic field generated by one or more additional sources operating within the same spaces. One may not generally conclude that active boundary surfaces operating with given reflection coefficient moduli in one space or with certain interacting sources necessarily produce the same energy-related effects in another space or with dissimilar interacting sources or source positions. Because active boundary surfaces are secondary sources that interact with other sources, relationships between their reflection coefficient moduli and energy-related effects in adjacent fields are not necessarily straightforward. Further analysis is required to develop additional relationships and more extensive understanding for three-dimensional systems.

[1] H. F. Olson and E. G. May, "Electronic sound absorber," J. Acoust. Soc. Am. **25**, 1130–1136 (1953).

[2] H. F. Olson, "Electronic control of noise, vibration, and reverberation," J. Acoust. Soc. Am. **28**, 966–972 (1956).

[3] D. Guicking and K. Karcher, "Active impedance control for one-dimensional sound," J. Vibr. Acoust. **106**, 393–396 (1984).

[4] D. Guicking and M. Albrecht, "An electret loudspeaker for active acoustic systems," J. Vibr. Acoust. **106**, 397–398 (1984).

[5] D. Guicking, K. Karcher, and M. Rollwage, "Coherent active methods for applications in room acoustics," J. Acoust. Soc. Am. **78**, 1426–1434 (1985).

[6] T. Yagi, A. Imai, and M. Konishi, "Variable sound absorption system using velocity and pressure feedback loudspeaker," Proceedings of the International Symposium on Active Control of Sound and Vibration, Tokyo, Japan, 1991, pp. 433–438.

[7] P. Darlington and G. C. Nicholson, "Theoretical and practical constraints on the implementation of active acoustic boundary elements," Proceedings of the Second International Congress on Recent Developments in Air- and Structure-Borne Sound and Vibration, Auburn, Alabama, USA, 1992, pp. 1011–1018.

[8] F. Orduña-Bustamante and P. A. Nelson, "An adaptive controller for the active absorption of sound," J. Acoust. Soc. Am. **91**, 2740–2747 (1992).

[9] C. R. Fuller, M. J. Bronzel, C. A. Gentry, and D. E. Whittington, "Control of sound radiation/reflection with adaptive foams," Proceedings of Noise-Control 94, Ft. Lauderdale, Florida, USA, 1994, pp. 429–436.

[10] R. L. Clark and D. G. Cole, "Active damping of enclosed sound fields through direct rate feedback control," J. Acoust. Soc. Am. **97**, 1710–1716 (1995).

[11] X. Meynial, "Active materials for applications in room acoustics," Proceedings of the Third International Conference on Intelligent Materials, Lyon, France, 1996, pp. 968–973.

[12] P. Darlington and M. R. Avis, "Noise control in resonant soundfields using active absorbers," Proceedings of Inter-Noise 96, Liverpool, England, 1996, pp. 1121–1126.

[13] P. Darlington, "Active boundary control of enclosed sound fields," Proceedings of Inter-Noise 96, Liverpool, England, 1996, pp. 1127–1132.

[14] T. W. Leishman, "Active control of sound transmission through partitions composed of discretely controlled modules," Ph.D. thesis, The Pennsylvania State University, University Park, Pennsylvania, USA, 2000.

[15] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1968), pp. 375–376.

[16] In this paper, **boldface type** indicates vector quantities. The symbol $\mathbf{e}_x$ denotes a unit vector in the direction of increasing x. Superimposed circumflex marks ($\wedge$) indicate complex frequency domain amplitudes of acoustic or mechanical signals.

[17] F. J. Fahy, *Sound Intensity*, 2nd ed. (E&FN Spon, London, 1995).

[18] L. G. Beatty, "Acoustic impedance in a rigid-walled cylindrical sound channel terminated at both ends with active transducers," J. Acoust. Soc. Am. **36**, 1081–1089 (1964).

[19] W. L. Rogers, "Experiments using an active duct termination to control acoustic impedance," ASHRAE Trans. **79**, Pt. 2, 167–171 (1973).

[20] A. R. D. Curtis, P. A. Nelson, S. J. Elliott, and A. J. Bullmore, "Active suppression of acoustic resonance," J. Acoust. Soc. Am. **81**, 624–631 (1987).

[21] P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic, London, 1992).

[22] W. S. Kim and M. G. Prasad, "Acoustic impedance of a finite length duct with dynamic termination," J. Acoust. Soc. Am. **91**, 3575–3578 (1992).

[23] D. Thenail, O. Lacour, M. A. Galland, and M. Furstoss, "The active control of wall impedance," Acustica **83**, 1039–1044 (1997).

[24] Notice that the velocity sense of piston 2 is opposite that of piston 1 and that of the semi-infinite tube termination. Both pistons in this model are assumed to radiate *into* the enclosed medium, creating an adjacent local fluid compression with positive piston displacement.

[25] C. H. Hansen and S. D. Snyder, *Active Control of Noise and Vibration* (E&FN Spon, London, 1997).

[26] S. D. Sommerfeldt and P. J. Nashif, "An adaptive filtered-x algorithm for energy-based active control," J. Acoust. Soc. Am. **96**, 300–306 (1994).

[27] J. W. Parkins, S. D. Sommerfeldt, and J. Tichy, "Narrowband and broadband active control in an enclosure using the acoustic energy density," J. Acoust. Soc. Am. **108**, 192–203 (2000).

[28] Although the reflection coefficient modulus is uniform along the length of the lossless plane wave tube, the representation $|R|_{x=L}$ is used for consistency with the complex $R|_{x=L}$ that must be evaluated at the termination.

[29] The primary and secondary sources used in the studies were ostensibly transducers of finite internal impedance. In one instance (Ref. 11), a layer of absorptive material was positioned in the tube near the primary source.

A state feedback electro-acoustic transducer for active control of acoustic impedance

Toshiya Samejima[a)]
Department of Acoustic Design, Kyushu Institute of Design, 4-9-1, Shiobaru, Minami-ku, Fukuoka 815-8540, Japan

(Received 7 March 2002; revised 14 November 2002; accepted 25 November 2002)

In this paper, a new control system in which the acoustic impedance of an electro-acoustic transducer diaphragm can be actively varied by modifying design parameters is presented and its effectiveness is theoretically investigated. The proposed control system is based on a state-space description of the control system derived from an electrical equivalent circuit of an electro-acoustic transducer to which a differentiating circuit is connected, and is designed using modern control theory. The optimal quadratic regulator is used in the control system design, with its quadratic performance index formulated for producing desired acoustic impedance. Computer simulations indicate that the acoustic impedance of the diaphragm can be significantly varied over a wide frequency range that includes the range below the resonance frequency of the electro-acoustic transducer. A computer model of the proposed control system is used to illustrate its application to semi-active noise control in a duct. It is demonstrated that the proposed control system provides substantial reductions in the noise radiating from the outlet of the duct, both in the stiffness control range and in the mass control range.
[DOI: 10.1121/1.1538197]

PACS numbers: 43.50.Ki, 43.40.Vn, 43.50.Gf [MRS]

I. INTRODUCTION

In general, it is difficult to control a sound field at low frequencies by conventional methods such as installing sound absorbers or changing room shape. Furthermore, such solutions are large and typically heavy, and there are often architectural design restrictions on their use. In such cases, the use of electro-acoustic transducers to actively control the sound field is a viable option. The development of technology for the active control of sound fields has been the focus of a number of studies.

To date, most of the research on the active control of sound fields has been centered around methods that are based on the inverse filtering of the transfer functions from sound sources to receiving points in a sound field, and simulating the desired transfer functions as strictly as possible. However, in some cases, it is sufficient to alter dynamic properties of a sound field appreciably.

Dynamic properties between a sound source and a receiving point in a sound field are significantly affected by the poles of the transfer function between them. The poles of a transfer function can be controlled by introducing a feedback loop to the system under study. Olson and May's ''electronic sound absorber'' [1,2] is one of the most notable methods of applying feedback control to the active control of sound fields. In their system, the sound pressure at a microphone collocated with a loudspeaker is fed back through an electro-acoustic transducer. By choosing an appropriate feedback gain, the local sound pressure near the microphone can be reduced. This control method is sometimes referred to as an acoustic virtual earth method.

Many investigations based on Olson and May's device have been published since then. Swinbanks proposed another configuration using two or three sources for active noise control in a duct, and verified its stability using the Nyquist stability criterion.[3] Ross presented a convenient algorithm to calculate a closed loop active noise controller.[4] Eghtesadi *et al.* developed a modified control system into which a compensating circuit was introduced to prevent the instability due to acoustic feedback.[5] Trinder *et al.* reintroduced the acoustic virtual earth method into active noise control in a duct.[6] Hong *et al.* proposed a tight-coupled tandem attenuator composed of a cascade of two simple monopole attenuators.[7] Nelson *et al.* described the basic principle of the feedback control approach used in the area of active control of sound.[8] In the principle, the transfer function from loudspeaker input to microphone output, called the secondary path, corresponds to the control plant in a general feedback control problem. If the frequency response of the secondary path is relatively flat and free from phase shift, then a simple feedback controller that is an inverting amplifier with a high gain can be adopted to attenuate the sound near the microphone with the stability of the closed loop system. However, in general, the frequency response of the secondary path is not flat and free from phase shift, because it contains the electro-acoustic response of the loudspeaker, the acoustic characteristics of the path between the loudspeaker and microphone, and the electro-acoustic response of the microphone. In that case, compensating filters can be introduced into the feedback controller to correct for the phase shift in the secondary path for the stability. Thus, the feedback controller designed for one sound field may not work well in another sound field. Nelson *et al.* also developed the method to implement the feedback controller using adaptive digital

a)Electronic mail: samejima@kyushu-id.ac.jp

filters.[9] In their method, the feedback control problem is transformed into a feedforward problem, and a part of the feedback controller is adapted using an estimate of the primary disturbance as a reference signal. Consequently, the feedback controller depends on not only the control plant response but also the statistical properties of the primary disturbance.

In the context of active control of acoustic impedance rather than active control of sound, several control systems based on Olson and May's system were proposed. Nishimura *et al.* proposed an "active acoustic treatment" [10] in which the acoustic impedance on the surface is actively controlled using a configuration similar to that of Olson and May. Yagi *et al.* developed an active sound absorption control system that feeds back not only the sound pressure near the loudspeaker but also the vibration velocity of the diaphragm in order to control the mechanical impedance of the loudspeaker diaphragm as seen from the sound field.[11] Furstoss *et al.* presented a control system involving direct active control of the acoustic impedance of a loudspeaker, via the simultaneous processing of both vibration velocity of the diaphragm and sound pressure in its close vicinity, although their experiment used a feedforward technique.[12] They also studied Olson and May's feedback control system for control of surface impedance by a hybrid active/passive method. A further theoretical analysis was given for the case that both frequency and incidence angle are allowed to vary. Lacour *et al.* reported experiments on active noise control in enclosures using the active impedance changes developed by Furstoss *et al.*[13] Okda *et al.* presented an active noise control system that uses a motional feedback loudspeaker and can control the internal impedance of an electro-acoustic transducer without microphones.[14] Lane *et al.* also applied a motional feedback technique to active control of sound fields. At first they developed a method to compensate loudspeakers in order to approximate constant volume velocity behavior.[15] Then, this compensated loudspeaker was used as an actuator in direct rate feedback control for active damping of enclosures, which required a constant volume velocity source in its theory.[16] These previous studies influenced by Olson and May's work utilized classical control theory, which is based on transfer function analysis.

Several studies on the application of modern control theory based on the state-space method for active control of sound fields have been published and have shown attractive results.[17–19] Compared with classical control theory, modern control theory has a number of advantages, as follows. (1) It is assured that the designed controller is causal because a control system is analyzed and designed in the time domain on the basis of the concept of state variables. (2) More effective control is expected because not only the outputs of a control object, but also the state variables that express the internal state of the control object are fed back for control inputs. (3) The controller is optimized by minimizing a cost function quantizing the objective of control, i.e., the concept of optimization is easily introduced.

In the present study, the control system proposed by Olson and May is reformulated using modern control theory based on the state-space description of a control object, i.e., an electro-acoustic transducer, and a new control system capable of controlling the acoustic impedance of the diaphragm is developed. The control system is constructed so as to allow the effectiveness of the control to be optimized in a wide frequency range. Since only an electro-acoustic transducer is treated as a plant, the design of the controller is independent of the sound field into which the control system is to be introduced. Only the dynamics of an electro-acoustic transducer are altered so that the acoustic impedance of its diaphragm becomes the desirable value for an acoustic objective. Such control is classified into the semiactive control scheme, to which little attention has been paid for sound field control. The theoretical formulation of the proposed control system is confirmed by computer simulations. In addition, the proposed control system is shown to be suitable for use as acoustic material with actively enhanced actions through its application to semiactive noise control in a duct.

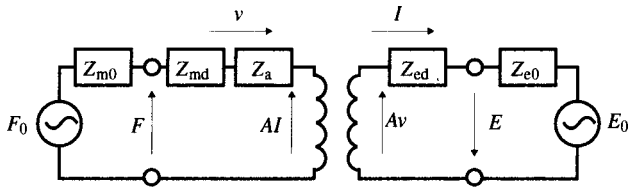


FIG. 1. Electrical equivalent circuit of an electro-dynamic electro-acoustic transducer.

II. STATE FEEDBACK CONTROL OF AN ELECTRO-ACOUSTIC TRANSDUCER

A. State-space description of an electro-acoustic transducer

Figure 1 shows the electrical equivalent circuit of an electro-dynamic electro-acoustic transducer, where E_0 is the electrical source voltage, I is the current, A is the force factor, Z_{e0} is the internal impedance of the power source, Z_{ed} is the electrical impedance of the voice coil, F_0 is the vibromotive force, v is the vibrating velocity of the diaphragm, Z_{m0} is the radiation impedance, Z_{md} is the mechanical impedance of the diaphragm, and Z_a is the acoustic impedance of the cabinet. Z_{e0}, Z_{ed}, and Z_{md} are given by

$$Z_{e0}=j\omega L_0+R_{e0}+\frac{1}{j\omega C_0},\quad Z_{\mathrm{ed}}=j\omega L_d+R_{\mathrm{ed}}+\frac{1}{j\omega C_d},$$
$$Z_{\mathrm{md}}=j\omega M_d+R_{\mathrm{md}}+\frac{K_d}{j\omega}, \tag{1}$$

where ω is the angular frequency, L_0, R_{e0}, and C_0 are the inductance, resistance, and capacitance component of Z_{e0}, respectively, L_d, R_{ed}, and C_d are the inductance, resistance, and capacitance of the voice coil respectively, M_d is the mass of the voice coil, R_{md} is the mechanical damping coefficient, and K_d is the mechanical stiffness of the suspension. Using the constants in Eq. (1) and assuming that Z_a consists only of the stiffness K_a, the equations of motion for the transducer are expressed as follows:

$$E_0-\left(L_0\frac{dI}{dt}+R_{e0}I+\frac{1}{C_0}\int I\,dt\right)$$
$$=\left(L_d\frac{dI}{dt}+R_{\mathrm{ed}}I+\frac{1}{C_d}\int I\,dt\right)-Av,$$
$$F=\left(M_d\frac{dv}{dt}+R_{\mathrm{md}}v+K_d\int v\,dt\right)+K_a\int v\,dt+AI. \tag{2}$$

Note that F represents the driving force given by pS_s, where p is the sound pressure in the vicinity of the diaphragm of the transducer, and S_s is the effective area of the diaphragm. Equation (2) can be transformed into a state-space description, as follows:

$$\dot{\mathbf{x}}=\mathbf{A}\mathbf{x}+\mathbf{b}_mF+\mathbf{b}_eE_0,\quad y=\mathbf{c}\mathbf{x}, \tag{3}$$

where

$$\mathbf{A}=\begin{bmatrix}0 & 1 & 0 & 0\\ -\dfrac{K_d+K_a}{M_d} & -\dfrac{R_{\mathrm{md}}}{M_d} & 0 & -\dfrac{A}{M_d}\\ 0 & 0 & 0 & 1\\ 0 & \dfrac{A}{L_0+L_d} & -\dfrac{1/C_0+1/C_d}{L_0+L_d} & -\dfrac{R_{e0}+R_{\mathrm{ed}}}{L_0+L_d}\end{bmatrix},$$

$$\mathbf{x}=\begin{Bmatrix}\int v\,dt\\ v\\ \int I\,dt\\ I\end{Bmatrix},\quad \mathbf{b}_m=\begin{Bmatrix}0\\ \dfrac{1}{M_d}\\ 0\\ 0\end{Bmatrix},\quad \mathbf{b}_e=\begin{Bmatrix}0\\ 0\\ 0\\ \dfrac{1}{L_0+L_d}\end{Bmatrix},$$

and y is an output of the system. The formulation of the output equation $y=\mathbf{c}\mathbf{x}$ depends on which state variable of the components of $\mathbf{x}$ is measured. If an extra coil is used to obtain the voltage proportional to the vibrating velocity v of the diaphragm, $\mathbf{c}$ is given by

$$\mathbf{c}=\{0\quad \mathbf{A}'\quad 0\quad 0\}, \tag{4}$$

where $\mathbf{A}'$ is the force factor of the extra coil.

Now, if state feedback such as $E_0=-\mathbf{f}\mathbf{x}$, where $\mathbf{f}$ is a state feedback gain vector, is employed, one obtains the following state-space description for a closed loop system:

$$\dot{\mathbf{x}}=(\mathbf{A}-\mathbf{b}_e\mathbf{f})\mathbf{x}+\mathbf{b}_mF,\quad y=\mathbf{c}\mathbf{x}. \tag{5}$$

If the system is controllable, the roots of the closed loop system can be determined arbitrarily by choosing an appropriate state feedback gain vector. Therefore, the characteristics of the electro-acoustic transducer can be controlled.

B. Control system design

For producing desired acoustic impedance at the transducer surface, we assume here that the objective is to reduce the error signal ε defined as

$$\varepsilon=p-Zv=F/S_s-Zv, \tag{6}$$

where Z corresponds to specific acoustic impedance that is set to a desired value by a user of the control system. The driving force F is given by

$$F=(K_d+K_a)\int v\,dt+R_{\mathrm{md}}v+M_d\dot{v}+AI. \tag{7}$$

Substituting Eq. (7) into Eq. (6) gives

$$\varepsilon=\left\{(K_d+K_a)\int v\,dt+R_{\mathrm{md}}v+M_d\dot{v}+AI\right\}\Big/S_s-Zv. \tag{8}$$

If the frequency range of interest is assumed to be the stiffness control range, then Eq. (8) can be approximated as follows:

$$\varepsilon\doteqdot\left\{(K_d+K_a)\int v\,dt+R_{\mathrm{md}}v+AI\right\}\Big/S_s-Zv$$
$$=\left\{\frac{K_d+K_a}{S_s}\;\frac{R_{\mathrm{md}}}{S_s}-Z\quad 0\quad \frac{A}{S_s}\right\}\mathbf{x}=\mathbf{r}_1\mathbf{x}. \tag{9}$$

In order to obtain a feedback controller that minimizes ε in the least squares sense, the optimal quadratic regulator can be employed. Its quadratic performance index is a time integral expressed as

$$J=\int_0^\infty(\mathbf{x}^T\mathbf{R}_1\mathbf{x}+E_0\mathbf{r}_2E_0)dt, \tag{10}$$

where $\mathbf{R}_1$ is a weighting matrix and r_2 is a weighting factor. The state feedback gain vector $\mathbf{f}$ that minimizes this performance index is calculated as follows:

$$\mathbf{f}=r_2^{-1}\mathbf{b}_e^T\mathbf{P}, \tag{11}$$

where $\mathbf{P}$ is the solution of the Riccati equation:

$$\mathbf{0}=\mathbf{P}\mathbf{A}+\mathbf{A}^T\mathbf{P}+\mathbf{R}_1-\mathbf{P}\mathbf{b}_er_2^{-1}\mathbf{b}_e^T\mathbf{P}. \tag{12}$$

Now, if $\mathbf{R}_1$ is given by

$$\mathbf{R}_1=\mathbf{r}_1^T\mathbf{r}_1, \tag{13}$$

then the term $\mathbf{x}\mathbf{R}_1\mathbf{x}^T$ in the integral of Eq. (10) is nearly equal to the square of ε. Thus, by adopting this means of state feedback control, in which the performance index (10) is minimized, the specific acoustic impedance of the diaphragm of the control system can be expected to take on a prescribed value Z as fast as possible when the diaphragm is exited by a disturbance. In addition, as this control system is designed in the time domain, there is no lack of causality regarding the realization of the feedback controller.

In the control system proposed by Olson and May, the sound pressure detected by the microphone, which is related to F, is fed back into the vibromotive force AI via the controller. The amplitude and phase characteristics of the controller are designed to minimize the sound pressure at the microphone. In order to obtain a feedback controller suitable for the same objective as that of Olson and May, Z is set to 0.

III. MODIFICATION OF THE CONTROL SYSTEM FOR A BROADBAND EFFECT

In the control system described in Sec. II, the inertia of the diaphragm was neglected when the performance index

was formulated. This is caused by the fact that the acceleration of the diaphragm is not involved in the state variables **x**. As a result of the neglected inertia, the frequency range in which the control system is effective is restricted to the stiffness control range. This feature is not critical when the control system is applied to noise control, because conventional passive methods such as sound absorbers and intervening barriers are relatively ineffective at low frequencies. However, if the control system can be made valid in the mass control range as well as in the stiffness control range, a system optimized over a wide band can be realized.

The inertia of the diaphragm neglected in Eq. (9) is expressed as

$$M_d\dot{v} = -(K_d+K_a)\int v\,dt - R_{md}v - AI + F. \tag{14}$$

Thus, if the driving force F can be measured, the inertia of the diaphragm can be controlled. F can be obtained by detecting the sound pressure near the diaphragm using a microphone. Feedback control systems using microphones and loudspeakers, such as Olson and May's system, appear to control the inertia of the diaphragm in this way, effectively reducing the sound pressure in the mass control range.

Here, a modified control system is presented in which a differentiating circuit is used to estimate the inertia of the diaphragm. A motional feedback loudspeaker to which a differentiating circuit is connected is regarded as a control object. A state-space description of this system is derived and the optimal quadratic regulator is applied to control the acoustic impedance of the diaphragm.

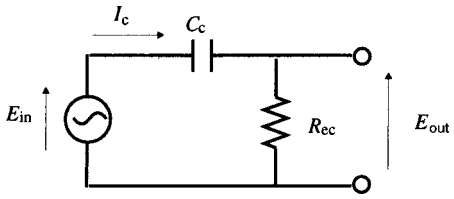


FIG. 2. Schematic of a differentiating circuit.

A. State-space description of an electro-acoustic transducer with a differentiating circuit

Figure 2 is a schematic of a differentiating circuit. If the output voltage of the extra coil of a motional feedback loudspeaker is given as the input voltage E_{in} of the differentiating circuit, the equation of motion for such an electrical circuit is expressed as

$$E_{\text{in}} = A'v = \Delta L_c\frac{dI_c}{dt} + R_{\text{ec}}I_c + \frac{1}{C_c}\int I_c\,dt, \tag{15}$$

where I_c is the current in the differentiating circuit, R_{ec} and C_c are the resistance and capacitance of the differentiating circuit, respectively, and ΔL_c is assumed to be sufficiently small. Equations (2) and (15) can be combined and transformed into a state-space description as expressed in Eq. (3). In this case, the matrix and the vectors in Eq. (3) are given by

$$\mathbf{A}=\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -\dfrac{K_d+K_a}{M_d} & -\dfrac{R_{\text{md}}}{M_d} & 0 & -\dfrac{A}{M_d} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \dfrac{A}{L_0+L_d} & -\dfrac{1/C_0+1/C_d}{L_0+L_d} & -\dfrac{R_{e0}+R_{\text{ed}}}{L_0+L_d} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & \dfrac{A'}{\Delta L_c} & 0 & 0 & -\dfrac{1/C_c}{\Delta L_c} & -\dfrac{R_{\text{ec}}}{\Delta L_c} \end{bmatrix},$$

$$\mathbf{x}=\begin{Bmatrix} \int v\,dt \\ v \\ \int I\,dt \\ I \\ \int I_c\,dt \\ I_c \end{Bmatrix}, \quad \mathbf{b}_m=\begin{Bmatrix} 0 \\ \dfrac{1}{M_d} \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix}, \quad \mathbf{b}_e=\begin{Bmatrix} 0 \\ 0 \\ 0 \\ \dfrac{1}{L_0+L_d} \\ 0 \\ 0 \end{Bmatrix}, \quad \mathbf{c}=\{0\ \ 0\ \ 0\ \ 0\ \ 0\ \ R_{\text{ec}}\},$$

where it is assumed that the output of the system is the output voltage E_{out} of the differentiating circuit.

B. Formulation of the performance index

The output voltage E_{out} of the differentiating circuit is given by

$$E_{\text{out}}=R_{\text{ec}}I_c=\frac{j\omega C_c R_{\text{ec}}}{1+j\omega C_c R_{\text{ec}}}A'v. \tag{16}$$

When $\omega C_c R_{\text{ec}}\ll 1$, Eq. (16) reduces to

$$R_{\text{ec}}I_c\doteqdot j\omega C_c R_{\text{ec}}A'v. \tag{17}$$

Thus, the acceleration $j\omega v$ of the diaphragm of the control system can be estimated from the current I_c in the differentiating circuit as follows:

$$j\omega v\doteqdot\frac{I_c}{C_c A'}. \tag{18}$$

Since I_c is a state variable of the system, the error signal ε can be expressed using the state variables $\mathbf{x}$ as follows:

$$\begin{aligned}\varepsilon&=\left\{(K_d+K_a)\int v\,dt+R_{\text{md}}v+M_d\dot{v}+AI\right\}\Big/ S_s-Z_v\\&\doteqdot\left\{(K_d+K_a)\int v\,dt+R_{\text{md}}v+\frac{M_d}{C_c A'}I_c+AI\right\}\Big/ S_s-Zv\\&=\left\{\frac{K_d+K_a}{S_s}\frac{R_{\text{md}}}{S_s}-Z\quad 0\quad\frac{A}{S_s}\quad 0\quad\frac{M_d}{C_c A' S_s}\right\}\mathbf{x}=\mathbf{r}_1'\mathbf{x}.\end{aligned} \tag{19}$$

The formulation of the weighting matrix in the performance index J,

$$\mathbf{R}_1=\mathbf{r}_1'^{T}\mathbf{r}_1', \tag{20}$$

therefore relates J to the square of ε in the mass control range as well as in the stiffness control range.

IV. CONFIGURATION OF THE CONTROL SYSTEM

Here, a practical method for realizing the feedback controller using a state estimator of a linear dynamic system is described, followed by an example configuration of the control system. If the system signal-to-noise ratios are sufficiently high, the system can be treated as deterministic and the state estimator can be designed as the Luenberger observer. However, if the signal-to-noise ratios are not very high, it is desirable that noise effect is taken into account in the design of the feedback controller and the state estimator is designed as the Kalman filter.[20] In either case, the state estimate $\mathbf{x}_e$ of the closed loop system represented by Eq. (5) is given by

$$\dot{\mathbf{x}}_e=(\mathbf{A}-\mathbf{b}_e\mathbf{f}-\mathbf{kc})\mathbf{x}_e+\mathbf{k}y, \tag{21}$$

where $\mathbf{k}$ is called the estimator gain vector. By carrying out $E_0=-\mathbf{f}\mathbf{x}_e$ instead of $E_0=-\mathbf{f}\mathbf{x}$, the transfer function $H(s)$ from y to the control command E_0 can be expressed as follows:

$$E_0(s)=-\mathbf{f}(s\mathbf{I}-\mathbf{A}+\mathbf{b}_e\mathbf{f}+\mathbf{kc})^{-1}\mathbf{k}Y(s)=H(s)Y(s), \tag{22}$$

where the Laplace transforms of y and E_0 are $Y(s)$ and $E_0(s)$, respectively. Thus, the feedback controller can be realized as a single-input, single-output filter. Figure 3 shows the block diagram of an electro-acoustic transducer to which the above-described feedback controller is introduced.

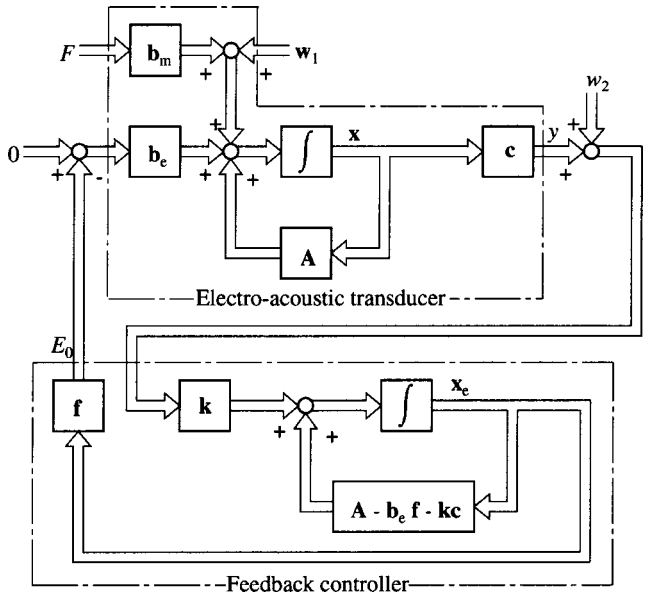


FIG. 3. Block diagram of an electro-acoustic transducer to which the feedback controller is introduced.

When the Kalman filter is used, $\mathbf{k}$ is determined by

$$\mathbf{k}=\mathbf{P}\mathbf{c}^T\mathbf{r}^{-1}. \tag{23}$$

In Eq. (23), $\mathbf{P}$ is the solution of the Riccati equation:

$$\mathbf{0}=(\mathbf{A}-\mathbf{b}_e\mathbf{f})\mathbf{P}+\mathbf{P}(\mathbf{A}-\mathbf{b}_e\mathbf{f})^T+\mathbf{Q}-\mathbf{P}\mathbf{c}^T r^{-1}\mathbf{c}\mathbf{P}, \tag{24}$$

where $\mathbf{Q}$ is the covariance matrix of the process noise vector $\mathbf{w}_1$ affecting the state variables of the system, and r is the variance of the observation noise w_2 affecting the output of the system. The well-known stochastic separation principle says that the control law of the form $E_0=-\mathbf{f}\mathbf{x}_e$ is optimal in the stochastic case when the Kalman filter is used; it will minimize the expected value of the performance index even if the system is very noisy.

Figure 4 shows the arrangement of the control system without the differentiating circuit, and Fig. 5 is for the con trol system with the differentiating circuit. In the arrangement of the control system with the differentiating circuit (Fig. 5), one can regard the differentiating circuit as a part of the feedback controller. If this is the case, the feedback con-

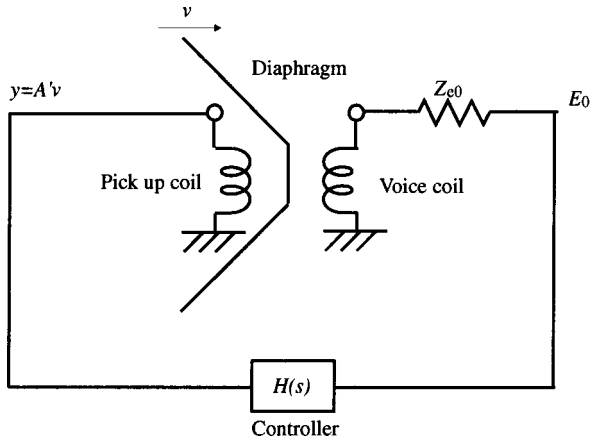


FIG. 4. Schematic of the proposed control system.

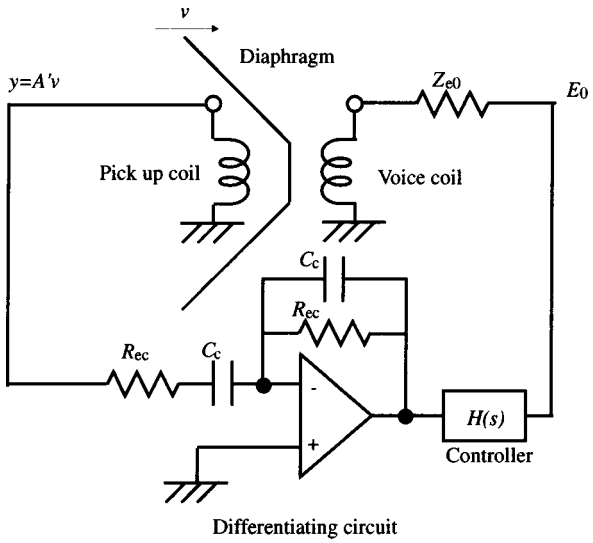


FIG. 5. Schematic of the proposed control system with a differentiating circuit.

troller can be implemented as $sH(s)$ with removing the differentiating circuit from the control system. Note that, when calculating $H(s)$ through the proposed procedure, it is necessary to treat a motional feedback loudspeaker to which a differentiating circuit is connected as a control plant. It is not until $H(s)$ is obtained that one can replace the differentiating circuit and $H(s)$ by $sH(s)$. This depends on the fact that one must utilize the quantity proportional to the acceleration of the diaphragm on the controller design through the state-space approach. Note also that, since the transfer function of a differentiating circuit is not exactly equal to s, this implementation of the feedback controller causes plant perturbations and may lead to instabilities of the system.

V. NUMERICAL STUDIES ON THE PROPOSED CONTROL SYSTEM

To assess the effectiveness of the proposed control system, computer simulations were performed on the theoretical basis presented in Secs. II and III. The objective of control was chosen as achieving zero acoustic impedance; the assigned impedance value is $Z=0$.

In general, when the optimal quadratic regulator theory is applied to practical control problems, it is difficult to find suitable values of the weighting matrix $\mathbf{R}_1$ and the weighting factor r_2 without actually calculating the response of the control system. This difficulty is one disadvantage of the optimal quadratic regulator theory. In Secs. II and III, $\mathbf{R}_1$ was formulated such that the performance index was related to the difference between the acoustic impedance of the diaphragm of the control system and its desired value. Thus, the design parameter in the computer simulations is only r_2.

A. Conditions for calculation

Table I gives the physical constants of the electro-acoustic transducer modeled in the computer simulations. Most of these constants were measured directly or estimated with sufficient accuracy.[21] Under the assumption that all the state variables were directly observable, the frequency responses of the vibrating velocity v of the diaphragm to the driving force F were calculated from the state equation of the control system represented in Eq. (5). The effectiveness of the state feedback control was evaluated in terms of the specific acoustic impedance $F/(vS_s)$ of the diaphragm seen from a sound field.

TABLE I. Physical constants of electro-acoustic transducer.

Force factor	$A=12.8\times10^{-4}$ Wb/m
Inductance of voice coil	$L_d=1\times10^{-3}$ H
Resistance of voice coil	$R_{ed}=6.7\ \Omega$
Capacitance of voice coil	$C_d=1\times10^{6}$ F
Mass of voice coil, cone, etc.	$M_d=2.8\times10^{-2}$ kg
Mechanical damping coefficient	$R_{md}=4.0$
Mechanical stiffness of suspension	$K_d=1400$ N/m
Effective area of diaphragm	$S_s=3.14\times10^{-2}$ m^2
Volume of cabinet	$V=3.5\times10^{-3}$ m^3
Force factor of pick up coil	$A'=12.8\times10^{-4}$ Wb/m
Inductance of differentiating circuit	$\Delta L_c=1\times10^{-4}$ H
Resistance of differentiating circuit	$R_{ec}=1\times10^{2}\ \Omega$
Capacitance of differentiating circuit	$C_c=1\times10^{-2}\ \mu$F

B. Results and discussion

Figure 6 shows the specific acoustic impedance of the control system without the differentiating circuit for variations in the weighting factor r_2. The weighting matrix $\mathbf{R}_1$ in the performance index is that of Eq. (13) neglecting the inertia of the diaphragm. Without control, i.e., with the electric terminal of the system open, the mechanical resonance of the electro-acoustic transducer causes a dip in the acoustic impedance. As r_2 decreases, the dip frequency becomes lower and the magnitude of the dip also changes. In the frequency range below approximately 130 Hz, the specific acoustic impedance of the diaphragm is reduced by introducing the control. In the frequency range considerably higher than the resonance frequency, i.e., in the mass control range, the control makes little difference to the specific acoustic imped-

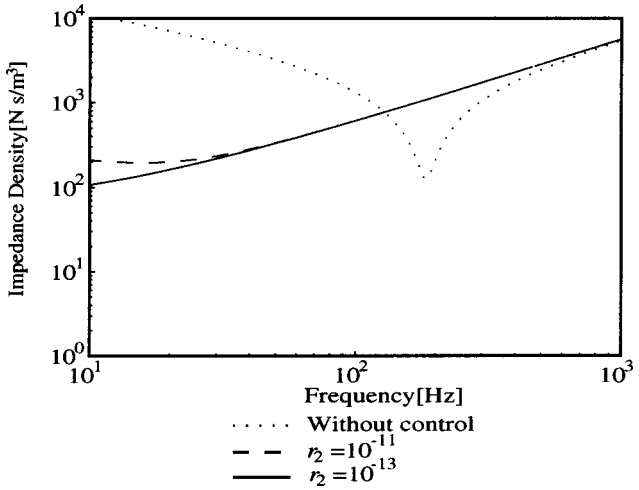


FIG. 6. Specific acoustic impedance of the control system without the differentiating circuit for variations in the weighting factor r_2.

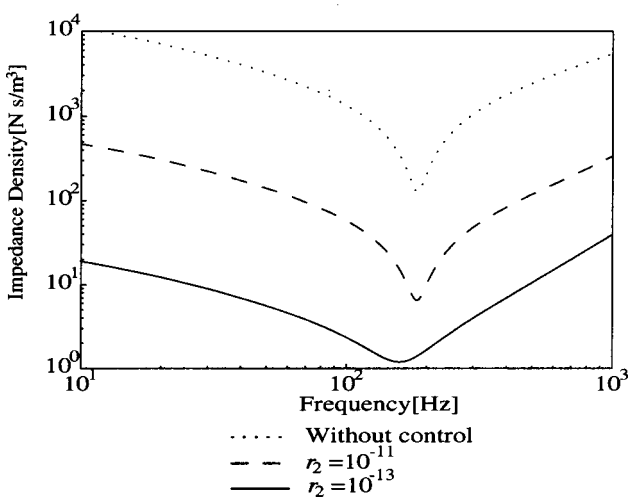


FIG. 7. Specific acoustic impedance of the control system with the differentiating circuit for variations in the weighting factor r_2.

ance. These results indicate that the electro-acoustic transducer is controlled so as to reduce the mechanical stiffness and mechanical resistance.

Figure 7 shows the specific acoustic impedance of the control system with the differentiating circuit. $\mathbf{R}_1$ of Eq. (20) was used, allowing for the inertia and elasticity of the diaphragm. In this case, as r_2 decreases, the magnitude of the dip is reduced while the dip frequency remains relatively constant. The specific acoustic impedance of the diaphragm is reduced by the introduction of this control in both the stiffness control range and the mass control range, indicating that all the components of mechanical impedance of the electro-acoustic transducer are reduced. As such, the acoustic impedance of the diaphragm of the control system is actively varied over a wide frequency range.

Although the theoretical formulation of the proposed control system was confirmed by the computer simulations, there are several practical issues that must be discussed for realizing the control system. One of those issues is related to the feedback gains and their practicality. The state feedback gain vector $\mathbf{f}$ involves the inverse of the weighting factor r_2, as expressed in Eq. (11). Thus, when r_2 is set to a very small value in order to achieve high control performance, the feedback gains, which refer to the absolute values of the components of $\mathbf{f}$, become very large. It is not easy to practically implement the control system involving the physically unreasonable gains. However, when the feedback controller is implemented using a state estimator of a linear dynamic system, the feedback gain has the frequency characteristics given by $(j\omega\mathbf{I}-\mathbf{A}+\mathbf{b}_e\mathbf{f}+\mathbf{kc})^{-1}$. This component may compensate for the large components of $\mathbf{f}$, leading to an acceptable feedback gain. Modeling errors and plant perturbations are other important issues. In general, it is difficult to model a plant accurately. In addition, perturbations of a plant may also occur due to variations in physical conditions such as temperature and humidity. These bring about deviations of the real plant from its nominal model, which is called plant uncertainties. The proposed feedback controller is designed for a nominal model of a plant. Thus, when this feedback controller is introduced into the real plant, plant uncertainties

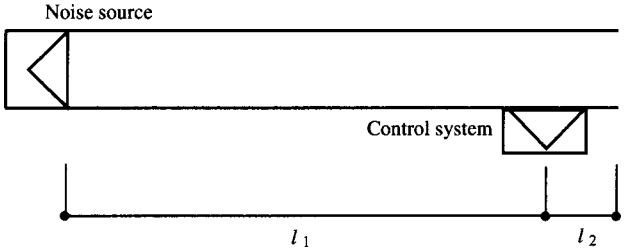


FIG. 8. Schematic of a duct equipped with the proposed control system for computer simulations.

construct an extra closed loop, which may cause instabilities of the whole closed loop system. If it is actually difficult to make the closed loop system stable, it is necessary to take into account plant uncertainties in the design of the feedback controller. This is accomplished by using robust control theory such as H_∞ control theory, which is an effective scheme for designing feedback controllers that accommodate both performance and stability in an optimal and robust manner. These practical aspects rather than theoretical ones will be explored in future work.

VI. APPLICATION TO SEMIACTIVE NOISE CONTROL IN A DUCT

Application of the proposed control system is demonstrated through computer simulations of semiactive noise control in a duct. The active control of one-dimensional acoustic waves propagating in a duct at low frequencies has become one of the classic problems in active noise control. However, there are some problems in the practical implementation of conventional active noise control systems using microphones and loudspeakers. This type of system has disadvantages under severe conditions such as extreme dustiness, high temperatures, or high humidity. In addition, the loudspeaker will not radiate sound efficiently in the frequency range lower than the resonance frequency of the loudspeaker. On the other hand, the only part of the proposed control system seen from a sound field is the diaphragm, and it behaves solely as a mechanical vibration system, dynamic properties of which are acoustically optimized even below the resonance frequency. Thus, it is expected that the installation of the proposed control system in a duct should prove simple and that the control system should run reliably in a wide frequency range.

A. Conditions for calculation

The proposed control system is mounted on the duct wall, as shown in Fig. 8. Assuming that the sound field in the duct is one-dimensional, and the effective diameter of the diaphragm of the control system is small compared to the wavelength, the duct can be treated as a transmission line. Figure 9 shows its equivalent circuit, where p_1 is the con stant pressure of the noise source, Z_s is the specific acoustic impedance of the diaphragm of the control system as seen from the sound field, S_s is the effective area of the diaphragm of the control system, S_d is the cross-sectional area of the duct, and l_1 and l_2 are the effective lengths between the noise source and the control system and between the

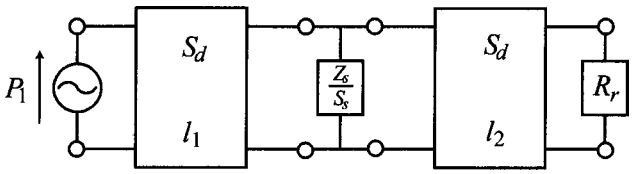


FIG. 9. Electrical equivalent circuit of a duct equipped with the proposed control system.

control system and the outlet of the duct, respectively; the circuit is loaded with the radiation resistance R_r of the outlet of the duct, including the radiation reactance as the correction for the effective length of the duct. If $R_r \ll \rho c$, the insertion loss IL due to the proposed control system can be given by

$$\mathrm{IL} = 20 \log \left| 1 + j \frac{S_s \rho c \sin kl_1 \sin kl_2}{S_d Z_s \sin k(l_1 + l_2)} \right|, \tag{25}$$

where ρ is the density of the medium, c is the speed of sound, and k is the wave number.[22]

In the computer simulations, the values of the specific acoustic impedance shown in Figs. 6 and 7 were used for Z_s, and the duct sizes were set as $l_1 = 1.645$ m, $l_2 = 0.185$ m, and $S_d = 0.29 \times 0.29$ m^2. The cut-on frequency of the first higher order mode was 594 Hz in this duct.

B. Results and discussion

Figure 10 shows the insertion loss IL due to the control system without the differentiating circuit for variations in the weighting factor r_2. A positive IL indicates that the sound power radiating from the outlet of the duct is reduced. Without control, i.e., with the electric terminal of the system open, sharp peaks due to the acoustic resonance of the duct occurring when $\sin k(l_1 + l_2) = 0$ are observed at equal intervals. The mechanical resonance of the electro-acoustic transducer also results in a sharp peak, because Z_s becomes minimal. This effect is similar to that brought about by a resonator muffler. The resonance frequency of the electro-acoustic transducer in this computer simulation is about 182 Hz. Thus, the peak at about 188 Hz is due to the mechanical

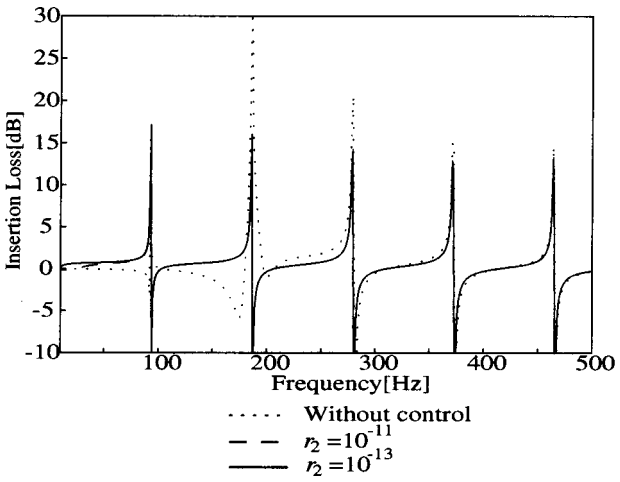


FIG. 10. Insertion loss IL due to control without the differentiating circuit for variations in the weighting factor r_2.

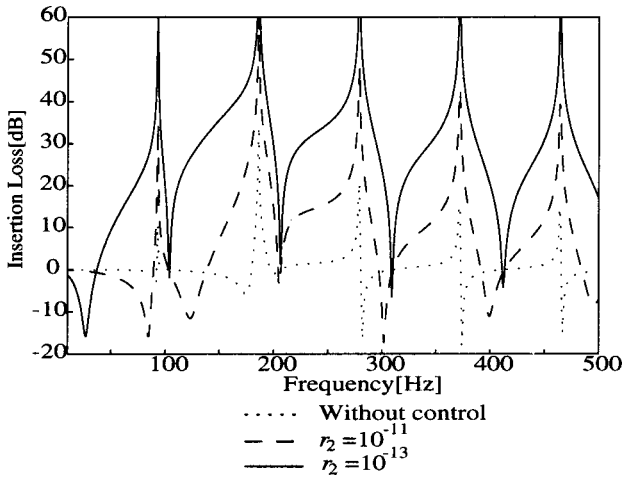


FIG. 11. Insertion loss IL due to control with the differentiating circuit for variations in the weighting factor r_2.

resonance of the electro-acoustic transducer as well as the acoustic resonance of the duct. In the frequency range lower than about 200 Hz, the control system brings about slight changes in the values of IL, attributable to changes in the specific acoustic impedance of the diaphragm of the control system below about 200 Hz as a result of the control, as shown in Fig. 6. However, as the reduction of the specific acoustic impedance is quite small, the values of IL do not become large.

Figure 11 shows the insertion loss IL due to the control system with the differentiating circuit for variations in the weighting factor r_2. As r_2 decreases, IL becomes more positive. When $r_2 = 10^{-13}$, a significant level of sound attenuation is achieved in a wide frequency range, i.e., both below and above the resonance frequency of the electro-acoustic transducer. This stems from the fact that the specific acoustic impedance of the diaphragm is significantly reduced by the introduction of this control in both the stiffness control range and mass control range, as shown in Fig. 7. While IL increases with the control of $r_2 = 10^{-13}$ over most of the frequency range above 40 Hz, it decreases below that frequency; a -15 dB dip occurs at about 25 Hz. Equation (25) indicates that this dip occurs when $1 + jS_s \rho c \sin kl_1 \sin kl_2 / \{S_d Z_s \sin k(l_1 + l_2)\}$ is nearly equal to 0. This can be avoided by adjusting each component of the mechanical impedance Z_{md} of the diaphragm or that of the assigned acoustic impedance Z. By calculating IL using Eq. (25) for variations in the specific acoustic impedance of the diaphragm Z_s, one can observe that, for example, a relatively large resistance component of Z_s does not bring about the dip.

VII. CONCLUSIONS

A control system in which the acoustic impedance of the system diaphragm is controlled was developed by recon structing Olson and May's control system and designing the controller using modern control theory. The optimal quadratic regulator was used in the control system design, and the quadratic performance index was formulated for producing desired acoustic impedance. As this design method is

formulated in the time domain, it is guaranteed that the designed controller is causal. Furthermore, the control system described herein can be realized without microphones. As such, the system can be easily installed and is expected to perform reliably and with high stability even under severe conditions.

Computer simulations confirmed that the proposed control system effectively reduces the acoustic impedance of the diaphragm over a wide frequency range, including the range below the resonance frequency of the electro-acoustic transducer. The proposed control system serves as an acoustic impedance surface, the impedance of which can be changed arbitrarily by modifying the design parameters of the control system. Thus, this system provides a means of tuning the characteristics of a sound field. As an application of the proposed control system, semiactive noise control in a duct was examined using a computer model. It was shown that the proposed control system provides noise reduction in both the stiffness control range and mass control range, which will not be achievable with conventional active noise control systems using microphones and loudspeakers.

ACKNOWLEDGMENT

The author would like to thank K. Komatsu for his assistance in the computer simulation.

[1] H. F. Olson and E. G. May, ''Electronic sound absorber,'' J. Acoust. Soc. Am. **25**, 1130–1136 (1953).
[2] H. F. Olson, ''Electronic control of noise, vibration, and reverberation,'' J. Acoust. Soc. Am. **28**, 966–972 (1956).
[3] M. A. Swinbanks, ''The active control of sound propagation in long ducts,'' J. Sound Vib. **27**, 411–436 (1973).
[4] C. F. Ross, ''An algorithm for designing a broadband active sound control system,'' J. Sound Vib. **80**, 373–380 (1982).
[5] Kh. Eghtesadi and H. G. Leventhall, ''Active attenuation of noise-The monopole system,'' J. Acoust. Soc. Am. **71**, 608–611 (1982).
[6] M. C. J. Trinder and P. A. Nelson, ''Active noise control in finite length ducts,'' J. Sound Vib. **89**, 95–105 (1983).
[7] W. K. W. Hong, Kh. Eghtesadi, and H. G. Leventhall, ''The tight-coupled monopole(TCM) and tight-coupled tandem(TCT) attenuators: Theoretical aspects and experimental attenuation in an air duct,'' J. Acoust. Soc. Am. **81**, 376–388 (1987).
[8] P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic, London, 1992), Chap. 7.
[9] S. J. Elliott, T. J. Sutton, B. Rafaely, and M. Johnson, ''Design of feedback controllers using a feedforward approach,'' Proceedings ACTIVE95, 1995, pp. 863–874.
[10] M. Nishimura, K. Ohnishi, W. P. Patrick, and A. C. Zander, ''Development of active acoustic treatment phase 1. Basic concept and development of AAT-cell,'' Proceedings ACTIVE97, 1997, pp. 319–330.
[11] T. Yagi, A. Imai, and M. Konishi, ''Variable sound absorption system using velocity and pressure feedback loudspeaker,'' Proceedings of the International Symposium on Active Control of Sound and Vibration, 1991, pp. 433–438.
[12] M. Furstoss, D. Thenail, and M. A. Galland, ''Surface impedance control for sound absorption: Direct and hybrid passive/active strategies,'' J. Sound Vib. **203**, 219–236 (1997).
[13] O. Lacour, M. A. Galland, and D. Thenail, ''Preliminary experiments on noise reduction in cavities using active impedance changes,'' J. Sound Vib. **230**, 69–99 (2000).
[14] J. Okda, T. Nishimura, Y. Nishimura, T. Usagawa, and M. Ebata, ''Active noise control system using motional feedback loudspeaker,'' J. Acoust. Soc. Jpn. (E) **12**, 291–297 (1991).
[15] S. A. Lane and R. L. Clark, ''Improving loudspeaker performance for active noise control applications,'' J. Audio Eng. Soc. **46**, 508–519 (1998).
[16] S. A. Lane and R. L. Clark, ''Dissipative feedback control of a reverberant enclosure using a constant volume velocity source,'' ASME J. Vibr. Acoust. **120**, 987–993 (1998).
[17] J. L. Dohner and R. Shoureshi, ''Modal control of acoustic plants,'' ASME J. Vib., Acoust., Stress, Reliab. Des. **111**, 326–330 (1989).
[18] M. R. Bai, C. Shieh, and D. Lee, ''Development of a modal LQG active noise canceler,'' Proceedings ACTIVE95, 1995, pp. 911–920.
[19] J. L. Dohner, and R. Shoureshi, ''A method for active noise control using a source-point model,'' J. Acoust. Soc. Am. **86**, 1053–1059 (1989).
[20] R. E. Kalman and R. S. Bucy, ''New results in linear filtering and prediction theory,'' ASME J. Basic Eng. **83**, 95–108 (1961).
[21] I. C. Shepherd, A. Cabelli, and R. F. Lafontaine, ''Characteristics of loudspeakers operating in an active noise attenuator,'' J. Sound Vib. **110**, 471–481 (1986).
[22] J. Okda, ''Calculation of the property of resonator type mufflers'' (in Japanese), J. Acoust. Soc. Jpn. **27**, 363–372 (1971).

Noise sensitivity and reactions to noise and other environmental conditions

Henk M. E. Miedema[a] and Henk Vos
TNO-Inro, P.O. Box 6041, 2600 JA Delft, The Netherlands

(Received 24 September 2001; revised 13 December 2002; accepted 23 December 2002)

This article integrates findings from the literature and new results regarding noise sensitivity. The new results are based on analyses of 28 combined datasets ($N=23\,038$), and separate analyses of a large aircraft noise study ($N=10\,939$). Three topics regarding noise sensitivity are discussed, namely, its relationship with noise exposure, its working mechanism, and the scope of its influence. (1) A previous review found that noise sensitivity has no relationship with noise exposure. The current analyses give consistent results, and show that there is at most a very weak positive relationship. (2) It was observed earlier that noise sensitivity alters the effect of noise exposure on noise annoyance, and does not (only) have an additive effect. The current analyses confirm this, and show that the relation of the annoyance score with the noise exposure is relatively flat for nonsensitives while it is steeper for sensitives. (3) Previous studies showed that noise sensitivity also influences reactions other than noise annoyance. The current analyses of the aircraft noise study extend these results, but also indicate that noise sensitivity has relatively little influence on reactions to nonenvironmental conditions.
[DOI: 10.1121/1.1547437]

PACS numbers: 43.50.Sr, 43.50.Qp [MRS]

I. INTRODUCTION

The percentage of persons annoyed by noise from transportation (aircraft, road traffic, railways) is related to the noise exposure level, and this relationship depends on the type of noise source (see Schultz, 1978; Kryter, 1982a,b; Fidell *et al.*, 1991; Miedema and Vos, 1998; Miedema and Oudshoorn, 2001). Research has shown that, in addition to the noise exposure, a variety of factors influence noise annoyance (Fields, 1993). One of these factors is noise sensitivity (e.g., see McKennel, 1963; Langdon, 1976; and see for reviews: Job, 1988, 1999; Fields, 1993). The correlation of noise sensitivity with subjective reactions to noise are high, in the order of 0.30 (see Job, 1988), so that it explains in the order of 9% of the variance in the reactions. Miedema and Vos (1998) found that the difference in noise annoyance between low and high noise sensitive persons (three categories: low, medium, high) was equal to the difference caused by an 11-dB change of the noise exposure (noise metric: DNL). Because of the large effect on annoyance, a better understanding of noise sensitivity is important. Simple exposure-response relationships alone do not provide sufficient understanding of the working mechanisms through which noise causes adverse effects, and the role of factors such as noise sensitivity therein. A better understanding of the working mechanisms will improve the scientific basis for efforts to reduce adverse noise effects. This article integrates findings from literature with new results, with a view to facilitating and stimulating the development of theories regarding noise sensitivity and its role in the mechanisms through which noise induces effects. Three specific objectives are presented in the next section.

[a]Electronic mail: HME.Miedema@inro.tno.nl

II. OBJECTIVES

The basis of the interest in noise sensitivity is the strong association between noise sensitivity and noise annoyance, while noise sensitivity has a weak or no relation with noise exposure (see, e.g., Job, 1988). Figure 1 illustrates the pattern of associations between noise exposure, noise sensitivity, and noise annoyance. Their different relations with noise exposure imply that noise sensitivity and noise annoyance are separate factors and not just different ways of measuring noise annoyance. Several aspects of Fig. 1, which are illustrated in Fig. 2, are investigated.

Objective 1 of this article is to test the absence of an association between noise exposure and noise sensitivity [see Fig. 2(a)]. While the association between noise exposure and noise annoyance, and the association between noise sensitivity and noise annoyance have been demonstrated in comprehensive meta-analyses (see the Introduction), there is not a similar comprehensive demonstration of the absence of an association between noise exposure and noise sensitivity. Objective 1 is to test the absence of this association in a comprehensive meta-analysis.

Previous analyses regarding the relation between noise sensitivity and noise annoyance mainly concerned the question whether an association exists and how strong it is. There has been little attention for the mechanism through which noise sensitivity exerts its influence on noise annoyance. Figure 2(b) illustrates two possible mechanisms. One possibility is that sensitivity has an independent effect on annoyance, which adds to the effect of the noise exposure. Another possibility is that noise sensitivity alters the effect of the noise exposure. In the first case high sensitivity is expected to increase noise annoyance irrespective of the noise exposure, while in the second case noise sensitivity ''amplifies'' the effect of the noise exposure and has no effect on annoyance

 0001-4966/2003/113(3)/1492/13/$19.00

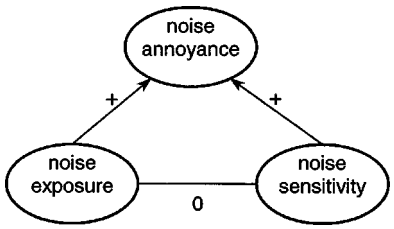


FIG. 1. A schematic overview of the relations between noise sensitivity, noise annoyance, and noise exposure. This pattern of interrelations will be further tested and elaborated in this article.

without noise exposure. *Objective 2* of this article is to investigate through which of these two mechanisms noise sensitivity acts.

The influence of noise sensitivity on noise annoyance has been demonstrated, but the scope of its influence may be wider. Noise annoyance is a response with a strong affective component to an environmental condition (noise). *Objective 3* of this article [see Fig. 2(c)] is to explore whether noise sensitivity also influences functional effects of noise (e.g., communication interference), whether it influences reactions to environmental conditions other than noise (e.g., odor), and whether it influences the evaluations of nonenvironmental conditions (e.g., access to facilities). If noise sensitivity has a negative impact on a wide range of affective responses, then it resembles the concept of ''negative affectivity.'' Persons with high negative affectivity experience discomfort in all circumstances, even in the absence of a clear agent that may cause discomfort (Watson and Clark, 1984; Watson and Pennebaker, 1989).

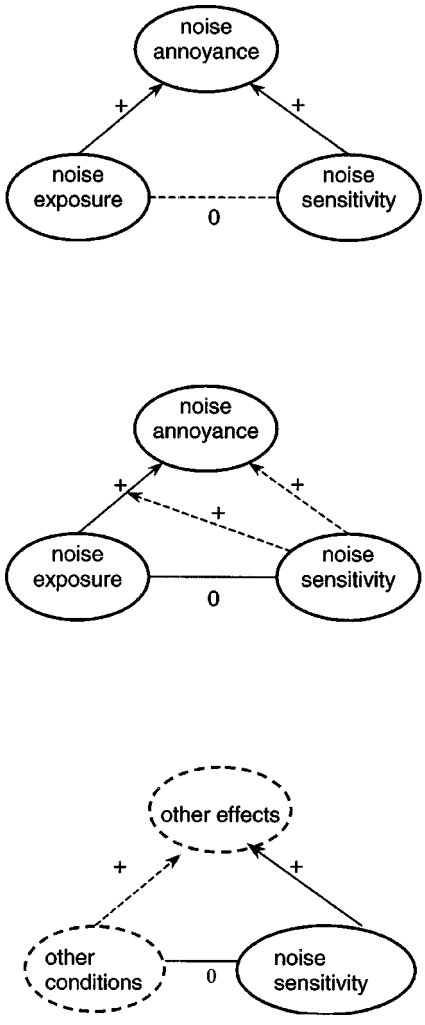


FIG. 2. As Fig. 1, but with the parts that are investigated in this article drawn in broken lines. The three figures correspond to the three objectives discussed in the text (top figure—objective 1, middle figure—objective 2, bottom figure—objective 3).

III. NOISE SENSITIVITY

A. Operational definition

Operational definitions of noise sensitivity are based on different types of questions. Respondents in social surveys (e.g., see McKennel, 1963; Langdon, 1976) or subjects in experiments (e.g., see Stansfeld, 1992; Ellermeier and Zimmer, 1997) are classified with respect to noise sensitivity on the basis of (1) self-reported characterizations of their noise sensitivity, (2) self-reported general attitudes towards noises, or (3) self-reported reactions to noise in specific situations. People are considered to be sensitive to noise if they have a high score on a direct noise sensitivity question, have a general negative attitude towards noises, or report strong reactions to specific noise situations. A physiological method of measuring noise sensitivity is not available, because there is no theory specifying the relation between a physiological measure and classifications on the basis of answers to noise sensitivity questions.

Many field surveys measure self-reported noise sensitivity with a single item of the above-mentioned type 1. A typical example of such a question is ''In general, how sensitive to noise are you?'', with options ''not at all,'' ''a little,'' ''moderately,'' ''considerably,'' and ''extremely.'' In general, a scale constructed from multiple items measuring the same attribute is more reliable than the response to a single item. Therefore, sets of noise sensitivity questions have been formulated, e.g., by Weinstein (1978), and, in German, by Zimmer and Ellermeier (1998). (Adapted versions of) Weinstein's noise sensitivity scale have been used relatively often. It consists of 21 items dealing with attitudes toward noise in general (questions of type 2), and emotional reactions to a variety of sounds (questions of type 3). Because it was originally designed specifically for a study with college students in the USA, Stansfeld (1992) adapted it for use in a general population in the UK.

B. Stability of noise sensitivity

Zimmer and Ellermeier (1999) found, as expected, that the test–retest reliability of Weinstein's and their own multiple item scales (0.87 and 0.91, respectively) was higher than the test–retest reliability for two single item scales (0.70 for a single item referring to *noise* with six response categories, and 0.83 for a single item referring to *sound* with ten response categories). Stability decreases over time. Zimmer and Ellermeier (1997) reported that the above test–retest reliabilities (0.70–0.91; Weinstein scale: 0.87) were found with a 4-week interval. Weinstein (1978) reported a test–retest reliability with a 9-week interval of 0.45. In a sample of depressed patients from four hospitals and their matched

controls, Stansfeld (1992: p. 21, Table 6) found test–retest reliabilities with a four months interval of 0.62–0.69. There was no difference in the stability of the noise sensitivity between the patients and the controls. In another sample of women living in a low or high aircraft noise exposure area, Stansfeld (1992) found test–retest reliabilities of 0.43–0.68 for 3-year intervals, again the correlation depending on the type of noise sensitivity scale. These latter results were based on retests of the low sensitive and the high sensitive individuals only. Stansfeld (1992, p. 14) reports that high sensitivity was more stable than low sensitivity.

C. Meaning of noise sensitivity

Zimmer and Ellermeier (1999) found that the correlations between Weinstein's multiple item scale, their own multiple item scale, and the two above-mentioned single item scales have values ranging from 0.57 to 0.79. The correlations between the two multiple item scales (0.79) and between the two single item scales (0.72) were higher than the correlations between the multiple and single item scales (0.57–0.64). This suggests that in particular the meaning of multiple item scales and the single item scales may not be fully the same. A further indication for a slightly different meaning of the multiple and single item scales comes from the pattern of correlations of these noise sensitivity scales with scales measuring depression, stress, state anxiety, trait anxiety, state anger, and trait anger. Zimmer and Ellermeier (1999) found that the Weinstein scale and the single items were related to the scales for depression, stress, and state anxiety. The main difference between them was that the Weinstein scale was also related to state anger, while the single items were not related significantly to state anger. All correlations of the Zimmer and Ellermeier scale were lower than the correlations of the Weinstein and the single item scales, except the correlation with state anger, which was between the correlation for the Weinstein scale and for the single items. The—slight—differences in the meaning of different noise sensitivity scales must be kept in mind when interpreting the outcomes of different studies. The field studies that are further analyzed here all contain a single item of the above-mentioned type 1 that measures self-reported noise sensitivity (see Sec. III A).

D. Noise sensitivity and neuroticism

Thomas and Jones (1988) and Stansfeld (1992) found positive correlations between neuroticism and noise sensitivity, which suggests that noise sensitivity may be an aspect of neuroticism. Evidence that noise sensitivity and neuroticism are different traits comes from a study that shows that they are correlated, but have a different correlation pattern with other variables. Belojevic *et al.* (1997) reported the correlations with seven aspects of sleep disturbance (e.g., difficulty falling asleep, tiredness after sleep) for both neuroticism and noise sensitivity. The correlation patterns are the same for neuroticism and for noise sensitivity. In addition, Belojevic and Jakovljevic (1997) reported correlations with the duration of opening the windows in the high noise zone: while neuroticism is not correlated with opening of the windows, noise sensitive persons open their windows less than nonsensitive persons. Dornic and Haaksonen (1989) also found different correlation patterns for noise sensitivity and neuroticism. They found that the noise levels chosen as ''clearly annoying'' by subjects in the laboratory were negatively correlated with noise sensitivity, while they were not related to neuroticism. Thus, noise sensitivity and neuroticism have different relations with opening of windows, and with annoyance ratings of sounds. Hence, they are (correlated but) different traits.

IV. NOISE EXPOSURE AND NOISE ANNOYANCE MEASURES

In the analyses of the field studies that will be presented, DNL is used as the noise exposure metric. This noise metric is defined in terms of the ''average'' levels during daytime and nighttime, and applies a 10-dB penalty to noise in the night:

$$\mathrm{DNL} = 10\,\lg[(\tfrac{15}{24})\cdot 10^{\mathrm{LD}/10} + (\tfrac{9}{24})\cdot 10^{(\mathrm{LN}+10)/10}].$$

Here LD and LN are the L_{Aeq} as defined in ISO 1996-2 (1987) for the day (7–22 h) and the night (22–7 h), respectively. The results presented here are to a large extent independent of the choice of this noise metric, e.g., the same results would have been obtained if the new European noise metric L_{den} (EU/Env, 2002) would have been used.

Annoyance questions in different studies do not use the same number of response categories. Some questions have only three response categories while others use as many as 11 categories. In order to obtain annoyance measures for different studies that are more comparable, different sets of response categories were translated into a scale ranging from 0 to 100. The translation is based on the assumption that a set of annoyance categories divides the range from 0 to 100 in equally spaced intervals and that the words or scores offered as answers are at the center of each category. Then the general rule that gives the category scores on a scale from 0 to 100 is; $\mathrm{score}_{\mathrm{category}\ i} = 100(i-\frac{1}{2})/m$. Here i is the rank number of the category and m is the total number of categories.

V. DATA

Noise sensitivity will be discussed on the basis of findings described in the literature and additional results presented here. The new results are based on the data in the TNO archive of original datasets from field studies on self-reported effects of environmental noise. These studies concern different modes of transportation (aircraft, road traffic, and railway), and were carried out in Europe, North America, Australia, and Japan. As far as possible, a common set of variables has been derived for all studies, which includes, among others, noise exposure measures and annoyance measures.

Table I lists the studies in the TNO archive in which noise sensitivity was assessed, in addition to DNL and noise annoyance (see Miedema and Vos, 1998, for details regarding the derivation of DNL and annoyance), and gives the noise sensitivity questions that are used. The studies all include the variables gender and age. Table I shows which

TABLE I. Datasets included in the analyses in this paper with the sensitivity question and response categories. The study codes in the first column refer to Fields' (2001) catalog. Cases are counted in the last column if valid noise exposure, annoyance, and noise sensitivity data are available. ''C'' in the one but last column indicates that a communication disturbance question is included in the dataset, and ''F'' indicates that a fear/worry question is included. Key references are given at the bottom of the table.

Fields code			*N*
	Aircraft		
CAN-168	In general, how sensitive are you to noise? not at all; a little; moderately; considerably; extremely.		629
NET-371	*To which extent are you sensitive to noise? not at all sensitive;...; very sensitive.*	*C,F*	*10 939*
NET-379	How sensitive are you to noise? not at all sensitive;...; very sensitive.	C,F	154
UKD-242	Would you say that you are more sensitive or less sensitive than other people are to noise? less; same; more.	C	1952
	Road traffic		
AUS-329	Are you sensitive to noise? not at all; little; moderately; rather; much; very much.	C	783
CAN-168	In general, how sensitive are you to noise? not at all; a little; moderately; considerably; extremely.		566
FRA-092	Are you sensitive to noise in particular? not at all sensitive; hardly sensitive; fairly sensitive; very sensitive.		879
FRA-364	In a general way, do you think that you are sensitive to noise? not at all; a little; very much.		844
GER-192	I am often very sensitive to noise. not true; little true; moderately true; rather true; very true.		1570
GER-373	How sensitive to noise in general are you? not at all; little; rather; much; very much.	C	428
GER-374	How sensitive to noise of all kinds are you? not at all sensitive; somewhat sensitive; rather sensitive; strongly sensitive; extraordinarily sensitive.	C	575
JPN-369	How would you describe your sensitivity to noise? not at all sensitive; a little sensitive; rather sensitive; very sensitive.		799
JPN-382	How would you describe your sensitivity to noise? not at all sensitive; a little sensitive; rather sensitive; very sensitive.	C	718
NET-106	Are you sensitive to noise? not at all sensitive; not sensitive; just not sensitive; just sensitive; sensitive; very sensitive.	C	378
NET-258	Are you sensitive to noise? not at all sensitive; not sensitive; just not sensitive; just sensitive; sensitive; very sensitive.	C	277
NET-276	Are you in general sensitive to noise? not sensitive at all;...; very sensitive.		696
NET-362	Are you sensitive to noise in general? not at all sensitive; not sensitive; hardly sensitive; somewhat sensitive; sensitive; very sensitive.		293
SWE-368	How would you describe your sensitivity to noise? not at all sensitive; a little sensitive; rather sensitive; very sensitive.	C	1201
TRK-367	In general, are you a person, sensitive to noise? not at all; sensitive; very much sensitive.	C	122
UKD-071	Which of these statements would you pick to describe yourself? Noise never bothers me at all; I am very sensitive to noise; Neither.		2022
UKD-072	Would you say you were more sensitive or less sensitive than other people to noise? less; same; more.	C,F	890
UKD-157	Which of the following statements would you pick to describe yourself? I am not at all sensitive to noise, it never bothers me at all; I am very sensitive to noise, it bothers me a lot; neither.		302
UKD-242	Would you say that you are more sensitive or less sensitive than other people are to noise? less; same; more.		398
	Railway		
GER-192	I am often very sensitive to noise. not true; little true; moderately true; rather true; very true.		1559
JPN-370	How would you describe your sensitivity to noise? not at all sensitive; a little sensitive; rather sensitive; very sensitive.	C	435
NET-153	To conclude, how sensitive are you to noise? Use this card to give your answer. not sensitive at all; not sensitive; just not sensitive; neutral; only just sensitive; sensitive; very sensitive.	C	600
NET-276	Are you in general sensitive to noise? not sensitive at all;...; very sensitive.	C	264
SWE-365	Generally, how would you describe your sensitivity to noise? not at all sensitive; not sensitive; fairly sensitive; very sensitive.		2705
UKD-116	Would you say you were more sensitive or less sensitive than other people to noise? less; same (volunteered); more.	C,F	999

AUS-329 Lercher, P. (**1992**). ''Auswirkungen des Strassenverkehrs auf Lebensqualität und Gesundheit (Effects of Road Traffic on Quality of Life and Health),'' Innsbruck, Austria, Abteilung Sozialmedizin, Universität Innsbruck.

CAN-168 Birnie, S. E., Hall, F. L., and Taylor, S. M. (**1980**). ''Community Response to Noise from a General Aviation Airport,'' Noise Control Eng. **15**(1), 37–45.

FRA-092 Vallet, M., Maurin, M., Page, M. A., and Pachiaudi, G. (**1978**). ''Annoyance from and Habituation to Road Traffic Noise from Urban Expressways,'' J. Sound Vib. **60**(3), 423–440.

FRA-364 Vallet, M., Vernet, I., Champelovier, P., and Maurin, M. (**1996**). ''A Road Traffic Noise Index for the Night Time,'' Inter-noise 96, pp. 2345–2350, Liverpool, England.

GER-192 Knall, V., and Schümer, R. (**1983**). ''The Differing Annoyance Levels of Rail and Road Traffic Noise,'' J. Sound Vib. **87**(2), 321–326.

GER-373 Kastka, J., Borsch-Galetke, E., Guski, R., Krauth, J., Paulsen, R., Schümer, R., and Oliva, C. (**1995**). ''Longitudinal Study on Aircraft Noise Effect at Dusseldorf Airport 1981–1993.'' Proceedings of the 15th International Congress on Acoustics, Trondheim, pp. 447–451.

JPN-369 Yano, T., Yamashita, T., and Izumi, K. (**1991**). ''Community Response to Road Traffic Noise in Kumamoto,'' J. Sound Vib. **151**(3), 487–495.

JPN-370 Yano, T., Yamashita, T., and Izumi, K. (**1997**). ''Comparison of Community Annoyance from Railway Noise Evaluated by Different Category Scales,'' J. Sound Vib. **205**(4), 505–511.

JPN-382 Murase, S., Sato, T., Yano, T., Björkman, M., Rylander, R., and Dankittikul, W. (**2000**). ''Comparison of Path Models of Road Traffic Noise Annoyance in Sweden, Japan, and Thailand.'' Proceedings of WESTPRACVII (Seventh Western Pacific Regional Acousitcs Conference). Kumamoto, pp. 939–942.

TABLE I. (*Continued.*)

NET-106	Bitter, C. (**1979**). ''Perception and Experience of Traffic Noise in a Residential District along a State Highway,'' Urban Ecol. **4**, 161–177.
NET-153	de Jong, R. G. (**1979**). ''A Dutch Study on Railway Traffic Noise,'' J. Sound Vib. **66**(3), 497–502.
NET-258	de Jong, R. G. (**1981**). ''Beleving van Geluidwerende Voorzieningen Langs Rijksweg 10'' (Evaluation of Noise Abatement Measures Alongside Highway 10), Geluid en Omgeving **4**(1), 16–18. Dokumentnr: 801. IMG-TNO, Delft. Geluid en Omgeving **4**[1], 16–18.
NET-276	Miedema, H. M. E., and Berg, R. (**1988**). ''Community Response to Tramway Noise,'' J. Sound Vib. **120**, 341–346.
NET-362	Ericsz, W. J., Noordam, A., and Schoonderbeek, W. (**1986**). ''Trollificering van buslijn 9 in Arnhem,'' Onderzoek naar de effecten van geluidhinder. Noise series GA-HR-12-1. Ministerie VROM, Leidschendam.
NET-371	TNO, RIVM. (**1998**). ''Self Reported Health, Risk Perception and Residential Satisfaction Around Schiphol Airport, Netherlands.''
NET-379	van Dongen, J. E. F., Steenbekkers, J. H. M., and Vos, H. (**1999**). ''De kwaliteit van de leefomgeving rond Groningen Airport Eelde,'' 99.031, Leiden, TNO-PG.
SWE-365	Öhrström, E., and Skånberg, A. B. (**1996**). ''A Field Survey on Effects of Exposure to Noise and Vibration from Railway Traffic, Part 1: Annoyance and Activity Disturbance Effects,'' J. Sound Vib. **193**, 39–47.
SWE-385	Murase, S., Sato, T., Yano, T., Björkman, M., Rylander, R., and Dankittikul, W. (**2000**). ''Comparison of Path Models of Road Traffic Noise Annoyance in Sweden, Japan, and Thailand.'' Proceedings of WESTPRACVII (Seventh Western Pacific Regional Acoustics Conference). Kumamoto, pp. 939–942.
TRK-367	Kurra, S., Tamer, N., and Rice, C. G. (**1995**). ''Environmental Noise Pollution Research Project Report,'' Istanbul, Technical University.
UKD-242	Atkins, C., Nurse, K., and Richmond, C. (**1982**). ''Aircraft Noise Index Study: Tabulations of the Responses to the Social Surveys,'' DR Communication 8312 (1984), Civil Aviation Authority, London.
UKD-071	Langdon, F. J., and Buller, I. B. (**1977**). ''Road Traffic Noise and Disturbance to Sleep,'' J. Sound Vib. **50**(1), 13–28.
UKD-072	Sando, F. D., and Batty, V. (**1975**). ''Road Traffic and the Environment,'' Social Trends **5**, 64–69.
UKD-116	Fields, J. M. (**1977**). ''Railway Noise Annoyance in Residential Areas: Current Findings and Suggestions For Future Research,'' J. Sound Vib. **51**(3), 343–351.
UKD-157	Langdon, F. J., and Griffiths, I. D. (**1982**). ''Subjective Effects of Traffic Noise Exposure, II: Comparisons of Noise Indices, Response Scales, and the Effects of Changes in Noise Levels,'' J. Sound Vib. **82**(2), 171–180.
UKD-242	Brooker, P. (**1983**). ''Public Reaction to Aircraft Noise: Recent U.K. Studies,'' Inter-noise 83, pp. 951–955.

studies also include questions concerning conversation disturbance by the noise source concerned, and fear or worry related to the noise source. These questions are used to investigate the scope of the influence of noise sensitivity.

In addition to these datasets, a large aircraft noise study (NET-371) is analyzed. A description of that study has been given by Miedema *et al.* (2000). This study includes, among others, the variables DNL (the corrected version described in the reference is used here), noise annoyance, noise sensitivity (see Table I), gender, and age. Furthermore, it includes questions regarding conversation disturbance, worry, and odor annoyance caused by air traffic, and questions about the (un-)pleasant aspects of the surroundings. These latter questions will be used to investigate the scope of the influence of noise sensitivity.

The percentage of highly annoyed persons found in NET-371 at a given noise exposure level was in the order of four times higher than the percentage found in other aircraft noise studies (cf. Miedema and Oudshoorn, 2001). Because of this unusually high annoyance and because the study would dominate the outcome with its 10 939 cases, NET-371 and the ''combined dataset,'' consisting of the other datasets from Table I, are analyzed separately.

Noise sensitivity as well as questions regarding conversation disturbance, fear/worry, and odor annoyance are transformed into 0–100 scales by the same rule used for noise annoyance (see Sec. IV).

VI. ANALYSES AND RESULTS

Relations are analyzed with standard multiple regression analysis, in which the independent variables are entered simultaneously. Previous exposure-response analyses found an effect of the type of transportation and a large study effect (see references in the Introduction). In order to take these effects into account, indicator variables S_i are defined for all but one dataset (S_i is equal to 1 for the cases from dataset i and 0 for all other cases). These indicator variables are included as independent variables in the regression analyses of the combined dataset. In this way the relationship with the dependent variable for the individual datasets can differ by a constant. There are no high correlations between the independent variables that could cause unstable solutions, and there are no reasons to expect important violations of the assumptions underlying inferences from the outcomes (normality, linearity, homoscedasticy; cf. Tabachnick and Fidell, 1983, p. 77). The only exception is a possible curvilinear effect of age (Miedema and Vos, 1998), which is accounted for by including also the square of age as a predictor.

A. Noise sensitivity and noise exposure

The relation of noise sensitivity (dependent variable) with noise exposure (DNL) is investigated with a multiple regression analysis, with gender, age, and the square of age as covariates. The model of which the parameters are estimated is as follows:

$$\text{Sensitivity} = a + \Sigma_i a_i \cdot S_i + b \cdot \text{DNL} + c \cdot \text{Gender} + d_1 \cdot \text{Age} + d_2 \cdot \text{Age}^2. \quad (1)$$

The parameter a is determined by the dataset for which there is no indicator variable S_i (each S_i, and hence $\Sigma_i a_i \cdot S_i$, is equal to zero for all the cases in that dataset), i.e., for that dataset a is the intercept of the linear relationship between DNL and annoyance. The parameters a_i are additive ''corrections'' to a so that $(a+a_i)$ is the intercept for dataset i. The term $\Sigma_i a_i \cdot S_i$ is omitted in the application of the model to the single dataset NET-371.

Table II gives the estimates of b, c, d_1, d_2, and their standard errors based on the combined dataset and NET-371,

TABLE II. Estimates (B) of the parameters in Eq. (1), the model for prediction of noise sensitivity (0–100), and their standard errors (s.e.). The 28 intercepts for the combined datasets are not shown because they are not relevant in the present context. All coefficients are significant at the 1% level.

	Combined dataset ($r=0.282$)		NET-371 ($r=0.110$)	
	B	s.e.	B	s.e.
Constant			12.7	4.3
DNL (45–75)	0.08	0.03	0.19	0.07
Gender (0,1)	3	0.3	3	0.5
Age (12–98)	0.92	0.06	0.82	0.09
Age^2	−0.0097	0.001	−0.008	0.001

respectively. The 28 estimates of the intercepts for the combined dataset (a and the a_i) are omitted because they are not of particular interest here. Table II shows that DNL has at most a weak positive effect on noise sensitivity. In addition, it shows that noise sensitivity is a curvilinear function of age. This relation is similar to the curvilinear relation between noise annoyance and age found by Miedema and Vos (1998).

In addition, Table III presents the simple correlations between DNL and noise sensitivity. These 29 correlations are low, 22 are not significantly different from 0 (1% level, two tailed), and the 7 significant correlations have a positive sign. This different analysis also shows that noise exposure has at most a weak positive influence on noise sensitivity.

TABLE III. Correlations between DNL (45–75 dB) and noise sensitivity (0–100) for the individual datasets. The correlation with an * are significant at the 1% level.

Fields code	r
Aircraft	
CAN-168	−0.005
NET-379	0.018
NET-371	0.031*
UKD-242	−0.072
Road traffic	
AUS-329	0.135*
CAN-168	0.022
FRA-092	0.115*
FRA-364	0.039
GER-192	0.068*
GER-373	0.008
GER-374	−0.033
JPN-369	0.050
JPN-382	0.190*
NET-106	0.073
NET-258	0.132
NET-276	−0.025
NET-362	0.083
SWE-368	0.074*
TRK-367	−0.064
UKD-071	−0.023
UKD-072	−0.051
UKD-157	−0.022
UKD-242	−0.030
Railway	
GER-192	−0.027
JPN-370	0.152*
NET-153	−0.013
NET-276	0.137
SWE-365	0.003
UKD-116	−0.050

B. Working mechanism of noise sensitivity

The working mechanism is investigated with a multiple regression analysis in which annoyance is the dependent variable, and noise exposure (DNL), noise sensitivity, and the product of these two variables are the predictors. If the product term contributes significantly to the prediction of annoyance, then noise sensitivity interacts with DNL, and it can be concluded that it influences the effect of the noise exposure. The model of which the parameters are estimated is as follows:

$$\text{Effect} = a + \Sigma_i a_i \cdot S_i + b \cdot \text{DNL} + c \cdot \text{Sensitivity} + d \cdot \text{Sensitivity} \cdot \text{DNL}. \quad (2)$$

The effect in this equation is annoyance. (For the analyses in the next subsections, the effect is conversation disturbance, fear/worry, or odor annoyance, respectively). Table IV (combined dataset) and Table V (NET-371) give the estimates of the parameters b, c, and d and their standard errors. It is found that the interaction of DNL with sensitivity contributes strongly to annoyance.

C. Scope of the influence of noise sensitivity

To investigate the scope of the influence of noise sensitivity, the same analysis conducted for annoyance (see above) is carried out for conversation disturbance (combined dataset and NET-371), fear/worry related to the noise source (combined dataset and NET-371), and for odor annoyance (NET-371 only). Table IV (combined dataset) and Table V (NET-371) give the results. Because a conversation disturbance question was only included in a subset of the combined dataset, the analysis for annoyance also has been carried out for this subset. By comparing the results of this analysis with the results for conversation disturbance, the use of different datasets as a possible explanation of differences can be excluded. Table IV also gives these results, and shows that the contribution of the product term Noise sensitivity·DNL to annoyance is stronger than its contribution to conversation disturbance. The coefficients of the interaction terms do not come into each others 99% confidence interval (for conversation disturbance 0.0079 ± 0.0026, for annoyance 0.0114 ± 0.0026).

For the three studies in the combined dataset that measured fear (NET-379, UKD-72, UKD-116), there is no effect of noise sensitivity or its product with DNL on fear. The coefficients for these terms are not significant at the 1% (or 5%) level.

The results for NET-371 in Table V concerning communication disturbance and annoyance are similar to the results in Table IV for the combined dataset, except that noise sensitivity now has very small nonsignificant coefficients. In addition, Table V shows that the product between sensitivity and DNL not only contributes to conversation disturbance and annoyance, but also (much stronger) to worry about the air traffic and to odor annoyance from the aircraft. Thus, it

TABLE IV. Estimates (B) of parameters in Eq. (2) (for the prediction of conversation disturbance and noise annoyance), and their standard errors (s.e.). The effects are measured on a 0–100 scale. The 28 intercepts for individual datasets are not shown because they are not relevant in the present context. The estimates are based on the combined dataset. All coefficients are significant at the 1% level.

	Conversation disturbance (c.d.)		Annoyance			
	Studies with c.d. ($r=0.525$)		Studies with c.d. ($r=0.516$)		All studies ($r=0.556$)	
	B	s.e.	B	s.e.	B	s.e.
DNL (45–75)	0.62	0.06	0.94	0.07	0.87	0.04
Noise sensitivity (0–100)	−0.37	0.07	−0.42	0.09	−0.508	0.05
Noise sensitivity*DNL	0.0079	0.001	0.0114	0.001	0.0124	0.001

appears that noise sensitivity also influences the effects of environmental factors other than noise (for which DNL is used as an indicator).

In order to get an indication of how far the generality of noise sensitivity goes, data from NET-371 concerning the pleasant and unpleasant aspects of the surroundings are analyzed. Table VI gives the average noise sensitivity scores of the respondents who selected an aspect, as well as the average noise sensitivity of those who did not select the aspect. Respondents could also select the statement that there are no pleasant aspects, or the statement that there are no unpleasant aspects. Only 1% selected the statement that there are no pleasant aspects, and these respondents are on average more sensitive to noise. A considerably larger part, namely 12%, selected the statement that there are no unpleasant aspects, and these respondents are on average less sensitive to noise.

VII. NOISE SENSITIVITY AND NOISE EXPOSURE

Noise sensitivity has previously been found to be independent of the noise exposure (Job, 1988, p. 997, Table VII with overview of correlations noise/sensitivity found in eight studies) and independent of changes in noise exposure (Raw and Griffith, 1988). The results in Tables II and III confirm this. Analyses of the combined dataset ($N=23\,038$) and the large aircraft noise study NET-371 ($N=10\,939$) show that noise sensitivity at most has a very weak, positive relationship with noise exposure. The influence of noise exposure on noise sensitivity is very small compared to the influence of noise sensitivity on, e.g., noise annoyance. The very weak association between noise sensitivity and noise exposure cannot explain the strong influence of noise sensitivity on annoyance. The association between noise exposure and noise sensitivity is sufficiently low to be neglected in causal models of noise effects.

This absence of an association has implications regarding the nature of noise sensitivity. By definition, noise sensitivity is a personal trait if it is invariant over different conditions and stable over time (with stability gradually decreasing as the time interval is longer). The above-discussed lack of association between noise exposure and noise sensitivity is an important indication that noise sensitivity is invariant over different conditions. Together with the relative stability over time (see Sec. III) this indicates that noise sensitivity is a personal trait and not only a temporary or situational state.

VIII. WORKING MECHANISM OF NOISE SENSITIVITY

Stansfeld (1992, pp. 14–15, Table 3) found that noise sensitivity has an effect on annoyance, but this effect was much larger and significant only for the respondents in the high noise exposure area. The effect of noise sensitivity on annoyance was not significant in an area with low noise exposure. A similar interaction between the influences of noise exposure and noise sensitivity on annoyance was found in an earlier investigation with a slightly larger sample, including the same respondents (Stansfeld, 1992, pp. 14–15, Table 3).

Miedema and Vos (1998) found a main effect of noise sensitivity on annoyance. They translated this into the equivalent DNL difference, i.e., the increase in DNL that gives a difference in annoyance that is the same as the difference found between low and high noise sensitive persons. However, they did not investigate interactions. Here the

TABLE V. Estimates (B) of parameters in Eq. (2) (for the prediction of conversation disturbance, annoyance, worry, and odor annoyance related to the air traffic) and their standard errors (s.e.). The effects are measured on a 0–100 scale. Estimates are based on study NET-371. All coefficients are significant at the 1% level, except the ones marked with *.

	Conversation ($r=0.494$)		Noise annoyance ($r=0.484$)		Worry ($r=0.435$)		Odor annoyance ($r=0.425$)	
	B	s.e.	B	s.e.	B	s.e.	B	s.e.
Constant	−118.4	7.1	−98.3	9.2	−87.7	10.4	−63.7	9.6
DNL (45–75)	2.88	0.13	2.41	0.16	1.96	0.18	1.35	0.17
Sensitivity (0–100)	−0.09*	0.14	0.03*	0.18	−1.15	0.21	−1.81	0.19
Noise sensitivity*DNL	0.0058*	0.002	0.0084	0.003	0.0281	0.004	0.0376	0.003

TABLE VI. The average noise sensitivity (0–100) for those who did not select a statement as applying to his/her surrounding (no), and those who did select the statement (yes). In addition, the percentages of respondents are given for both categories. The same information is given between parentheses for respondents who did not select the final statements (there are no pleasant aspects; there are no unpleasant aspects). The data are from NET-371.

Pleasant aspects	No	Yes	Unpleasant aspects	No	Yes
Quiet	47(48)	42(44)	Not quiet	41(43)	54(54)
	63%(65%)	37%(35%)		72%(69%)	28%(31%)
Little traffic	46(47)	44(45)	Much traffic	43(46)	51(51)
	61%(62%)	39%(38%)		79%(77%)	21%(23%)
Space/space for living	46(47)	44(46)	Little space/space for living	45(46)	49(49)
	59%(60%)	41%(40%)		92%(91%)	8%(9%)
Stores nearby	45(47)	45(46)	Stores far away	45(47)	46(46)
	52%(53%)	48%(47%)		88%(86%)	12%(14%)
Work nearby	45(47)	44(46)	Work far away	45(47)	47(46)
	71%(71%)	29%(29%)		94%(94%)	6%(6%)
School nearby	45(47)	43(45)	School far away	45(47)	47(47)
	78%(78%)	22%(22%)		96%(96%)	4%(4%)
Center nearby/central	45(47)	44(46)	Center far away/not a	45(47)	47(47)
location	75%(75%)	25%(25%)	central location	92%(92%)	8%(8%)
Close to/good	45(47)	45(47)	Town far away	45(47)	48(47)
connections with town	56%(56%)	44%(44%)		95%(94%)	5%(6%)
Much green/possi-	45(46)	46(48)	Too little green/nature/	45(47)	48(48)
bilities for recreation	62%(63%)	38%(37%)	possibilities for recreation	90%(89%)	10%(11%)
Nice neighbor-	46(48)	43(45)	No nice neighbor-	44(46)	50(50)
hood/neighbors	60%(61%)	40%(39%)	hood/neighbors	92%(91%)	8%(9%)
Space/nice view	45(47)	45(47)	No nice view	45(46)	48(48)
	67%(67%)	33%(33%)		87%(86%)	13%(14%)
Nice dwelling	46(47)	45(46)	No nice dwelling	45(46)	50(50)
(garden)	43%(44%)	57%(56%)	(garden)	94%(93%)	6%(7%)
No environmental	46(47)	39(42)	Environmental pollution	43(45)	52(52)
pollution	87%(89%)	13%(11%)		77%(74%)	23%(26%)
Safe living environment	46(47)	42(44)	Unsafe living environment	44(46)	52(52)
	78%(79%)	22%(21%)		88%(87%)	12%(13%)
No pleasant aspects	99%	1%	No unpleasant aspects	88%	12%

same datasets combined with ten additional datasets are reanalyzed, and the data from the large aircraft noise study NET-371 is analyzed, now allowing for interaction with the noise exposure. Then a (strong) interaction between noise sensitivity and noise exposure is found (see Tables IV and V). Figures 3 and 4 illustrate this. They are obtained by substituting the parameter estimates from Tables IV and V in Eqs. (1) and (2). For the combined dataset, all S_i are set equal to 0 so that the intercept of the lines (a) is based on 1 of the 28 datasets. Consequently, the absolute ''height'' of the lines in Fig. 3 is more or less arbitrary. Sensitivity in Eqs. (1) and (2) has been set equal to 0 for the lines labeled sensi=0, and for the lines sensi=100 it has been set equal to 100. Thus, the difference between the lines sensi=0 and sensi=100 represents the upper boundary of the influence of noise sensitivity.

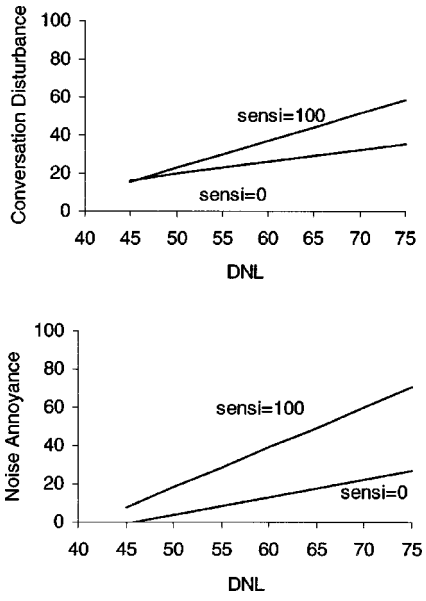


FIG. 3. Relations between, on the one hand, conversation disturbance or annoyance, and, on the other hand, DNL, when sensitivity is set equal to 0 or 100 in Eq. (2) with the parameter estimates from Table III (combined dataset).

With annoyance as the effect, the coefficient of the interaction term found on the basis of NET-371 (0.0084) is lower than this coefficient for the combined dataset (0.0124), but they are in the same order of magnitude (see Tables III and IV). These coefficients determine the differences in the slopes of the annoyance functions for sensi=0 and sensi =100, in Figs. 3 and 4. It is noteworthy that the lines for low and high sensitives converge to a low level of annoyance at low exposure levels (40–45 dB) in Fig. 3 (based on the combined dataset) but not in Fig. 4 (based on NET-371).

IX. SCOPE OF THE INFLUENCE OF NOISE SENSITIVITY

The influence of noise sensitivity on the perception of sound, disturbance of recall, or conversation by noise, noise annoyance, fear related to the noise source, physiological effects, sleep disturbance, and the evaluation of environmen-

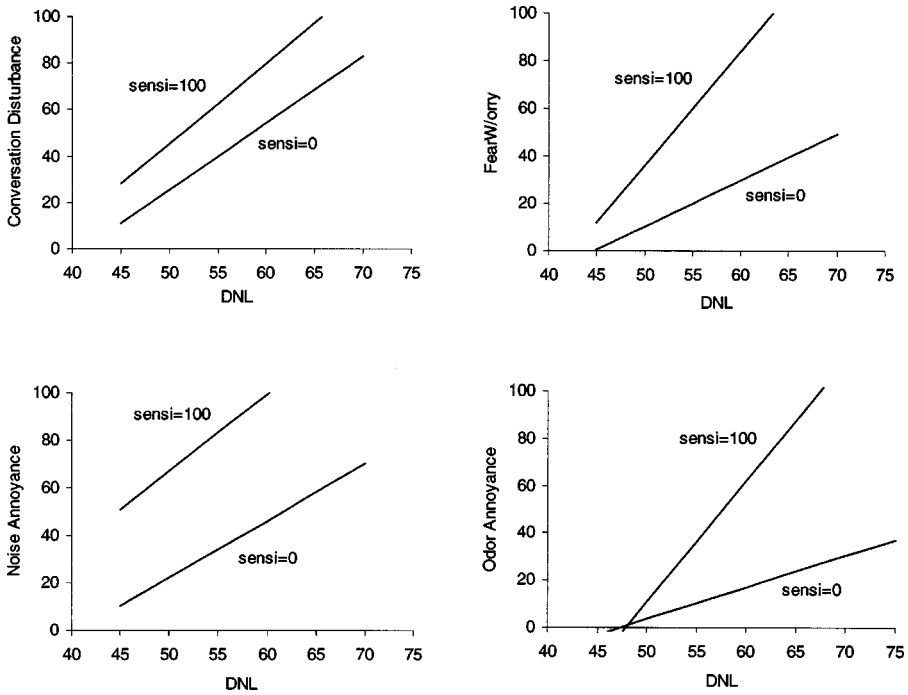


FIG. 4. Relations between, on the one hand, conversation disturbance, annoyance, fear/worry or odor annoyance, and, on the other hand, DNL, when sensitivity is set equal to 0 or 100 in Eq. (2) with parameter estimates from Table IV (NET-371).

tal factors in general is discussed in the following separate subsections. In the final subsection, limits to the scope of the influence of noise sensitivity are discussed.

A. Perception

Noise sensitive persons do not rate noise as being louder (Stephens, 1970; Waddell and Gronwall, 1984; Winneke and Neuf, 1992). Moreira and Bryan (1972) found a weak relation with the slope of the loudness function, but this is difficult to interpret without knowledge of the intercept. Ellermeier *et al.*, (2001) found no difference in absolute threshold, intensity discrimination, simple auditory reaction time, or growth rate of the loudness functions between low and high sensitive persons. Differences were found when the judgments involved evaluation of the noise, such as ratings of the unpleasantness or determinations of the uncomfortable loudness level (Thomas and Jones, 1988; Winneke and Neuf, 1992; Ellermeier *et al.,* 2001). Thus, while in the laboratory negative qualities of sounds are rated higher by the sensitive persons, there appears to be no difference in the pure perception of the sounds.

B. Disturbance (of recall or conversation) and annoyance

Ellermeier and Zimmer (1997) studied the effect of ''irrelevant speech'' on recall. They found that noise sensitivity was only weakly associated with recall performance, the strength of the association depending on the type of scale. Only one of the four noise sensitivity scales (see Sec. III), namely their own scale, had a significant relation with recall performance.

Stansfeld (1992) found no relation between hearing loss and noise sensitivity in a study with depressed patients and matched controls, although in general (neuro-sensory) hearing loss is associated with a larger detrimental effect of noise on speech comprehension. This finding suggests that vulnerability of speech comprehension to noise does not lead to higher self-reported noise sensitivity.

Figure 3, based on the current analyses of the combined dataset, illustrates that noise sensitivity has an effect on conversation disturbance, but has a stronger effect on noise annoyance. In Fig. 4, based on the analysis of NET-371, the difference in the slopes of the two lines for conversation disturbance and the difference in the slopes for annoyance are smaller. The effect on conversation disturbance presumably occurs because the rating of conversation disturbance not only reflects the frequency of occurrence of conversation disturbance, but also the evaluation of noise causing this disturbance. Noise annoyance has a stronger evaluative component and is found to be influenced more strongly by noise sensitivity.

The above findings suggest that the noise sensitivity does not so much influence actual interferences by noise, but has an influence on the evaluation of interferences. However, due to the limited and somewhat ambiguous evidence regarding recall, and the confounding of the self-reported speech interference with feelings regarding noise causing this interference, the evidence for this hypothesis is considered to be limited and inconclusive.

C. Anxiety

The relationship between noise sensitivity and (trait) anxiety has been documented in the literature (Zimmer and Ellermeier, 1998). The results in Table V and Fig. 4 indicate that noise sensitivity also has an important role in the generation of anxiety or worry by aircraft (state anxiety). The figure shows a similar pattern as found on the basis of the combined dataset for speech disturbance and annoyance (see Fig. 3), but with a larger divergence between low and high sensitives. Since fear/worry has a stronger affective/

emotional component, this is consistent with the hypothesis that especially affective/emotional reactions are influenced by noise sensitivity.

NET-379, a small study around a local airport that only accommodates small aircraft, the road traffic study UKD-72, and the railway study UKD-116 give a different outcome with respect to fear. Here no effect of noise sensitivity or the product term Noise sensitivity · DNL is found. The absence of an effect on fear in these studies may be explained by the low prevalence of fear, and a different nature of the fear. Fear around a small airport, roads, and railways may be different from the fear in NET-371, which is a busy international airport with large aircraft flying over. The noise sensitivity questions used in UKD-72 and UKD-116 were dichotomous, which also may have contributed to the lack of an effect.

D. Physiological effects

Stansfeld (1992) studied the effect of noise level and noise sensitivity (using two different noise sensitivity scales) on skin conductance response and on heart rate response in 18 depressed patients and 18 matched controls. For skin conductance, he found a small main effect of noise sensitivity, while a strong interaction between noise level and sensitivity was found: differences in responses between low and high sensitive persons were higher when the noise exposure was higher (Stansfeld, 1992, p. 27 and Fig. 3). This effect on skin conductance was only found with one of the noise sensitivity scales, while the following effect on heartbeat was only found with the other scale. For heart beat it was found that low levels of noise had no effect or induced a deceleration of the heart beat in both sensitive and nonsensitive persons, while higher noise levels induced accelerations, in noise sensitive persons only (Stansfeld, 1992, pp. 29–30 and Fig. 6). This is consistent with findings reported by Öhrström *et al.* (1988, p. 444), who reported a somewhat larger increase in heart rate after noise events during sleep for noise sensitive persons compared to nonsensitive persons.

E. Sleep disturbance

Lercher (1995, Fig. 1) found that sleep disturbance by road traffic noise was not related to noise sensitivity at low noise levels, while at higher noise levels the percentage of noise sensitive persons reporting sleep disturbance was much higher than the percentage nonsensitive persons reporting sleep disturbance. This is consistent with findings reported by Öhrström *et al.* (1988, p. 445), who report a larger effect of noise on time to fall asleep and on subjective sleep quality for noise sensitives compared to nonsensitives.

F. Odor annoyance

Winneke and Neuf (1992) compared the reactions of persons who are highly annoyed by noise in their residential area with the reactions of persons from the same areas who were little annoyed. They found that in the laboratory the highly annoyed persons were not only more annoyed by exposure to noise, but also by exposure to odor (H_2S) and exposure to environmental tobacco smoke. The same experiments were carried out with persons who were living close to an industrial odor source and who were highly annoyed by the odor and persons from the same area who were not annoyed by the odor. For these persons the same reaction pattern was found. In the laboratory the highly annoyed persons were not only more annoyed by exposure to odor (H_2S), but also by exposure to noise and to environmental tobacco smoke. This suggests that the same sensitivity contributes to high noise annoyance, high odor annoyance, and high annoyance by environmental tobacco smoke, and it supports the hypothesis that noise sensitivity also affects reactions to environmental factors other than noise. Results from the large aircraft noise study NET-371 in Table V and the corresponding Fig. 4 further demonstrate that noise sensitivity also has a strong effect on odor annoyance. The relationship between noise sensitivity and odor annoyance is similar to the relationship between noise sensitivity and self-reported noise effects found on the basis of the combined dataset, and similar to the relationship between noise sensitivity and worry found for NET-371.

G. Sensitivity to environmental factors in general

Langdon (1976) found that noise sensitivity has a strong relation with general environmental dissatisfaction. It is not likely that this relation between noise sensitivity and general dissatisfaction with the environment is mediated by higher noise annoyance, because noise annoyance was only very weakly related to general environmental dissatisfaction. It appears that a noise sensitive person is more concerned about the existence of environmental problems, and that this concern is not a consequence of a higher annoyance caused by noise pollution. Consistent findings have been reported by Meijer *et al.* (1985), who found that noise sensitive persons have less appreciation of their living environment.

H. Limits to the influence of noise sensitivity

In the previous subsections it has been found that noise sensitivity does not only influence effects of noise, but also reactions to other environmental factors. Here the limits to the influence of noise sensitivity are explored. Specifically, the hypothesis that noise sensitivity is an aspect of negative affectivity (experiencing discomfort under all circumstances) is discussed.

Broadbent (1972) reported that noise sensitive subjects were less favorable in their evaluation of food, holidays, and beauty. Thus, they showed a general negativity consistent with the hypothesis that noise sensitivity is an aspect of a general disposition to experiencing discomfort. This finding appears to indicate that noise sensitivity is an aspect of general negative affectivity. However, the evidence is ambiguous. The (permanent) influence of noise (and other environmental factors) may bring noise sensitive persons in a state which resembles negative affectivity, while it does not induce such a state in nonsensitive persons. This would cause correlations of noise sensitivity with aspects of negative affectivity, while noise sensitivity is not an aspect of negative affectivity.

Results from analyses discussed in previous subsections

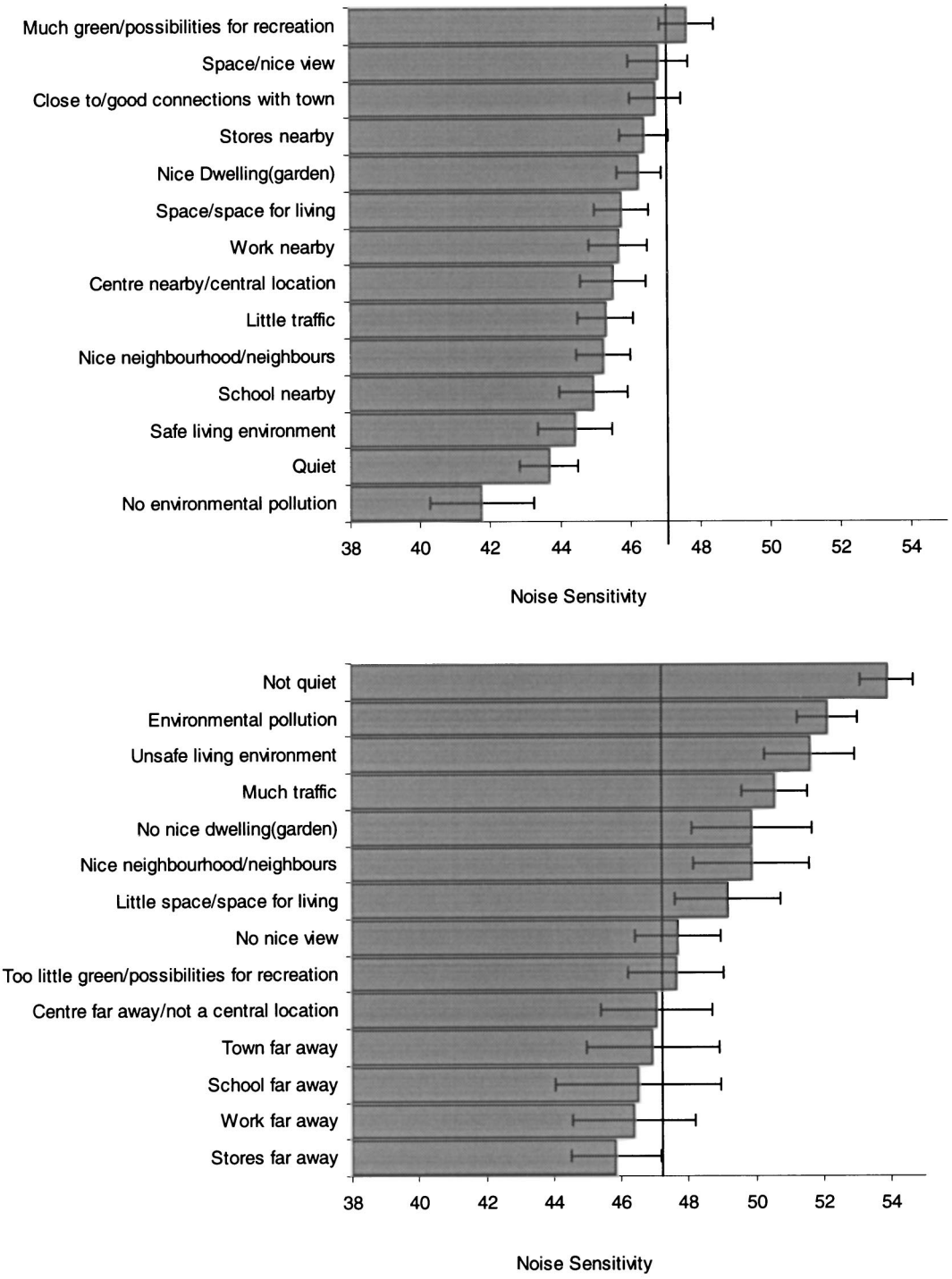


FIG. 5. Average noise sensitivity score for respondents who selected an aspect as applying to the surroundings, with a reference line for the overall mean noise sensitivity, and the 95% confidence intervals of the averages. (Results from Table VI, for those who did not select the final statements; NET-371.)

provide evidence for the hypothesis that noise sensitivity and negative affectivity are different traits. Persons with high negative affectivity experience discomfort under all circumstances, even in the absence of a clear agent that may cause discomfort (Watson and Clark, 1984; Watson and Pennebaker, 1989). Thus, if noise sensitivity is an aspect of negative affectivity, then the evaluation of noise by sensitive persons would be shifted towards the negative side independent of the noise exposure. This, however, is not what has been found (see Figs. 3 and 4). The outcome that noise sensitivity exerts its influence by changing the influence of noise contradicts the hypothesis that noise sensitivity is an aspect of a general negative affectivity.

Furthermore, Job *et al.* (1999) found that the correlations between self-ratings of sensitivity to aircraft noise, sensitivity to traffic noise, general sensitivity to loud noises, and general sensitivity to quiet noises were significant, while ratings of neighborhood parks and playgrounds did not correlate significantly with any of these sensitivity ratings. Thus noise sensitivity does not influence the evaluation of all cir-

cumstances, which indicates that it is not the same as general negative affectivity.

To further investigate whether noise sensitivity is an aspect of negative affectivity, data from NET-371 concerning the selection of pleasant and unpleasant aspects of the surroundings are used. Figure 5 illustrates the outcomes presented in Table VI, and gives for each aspect the average noise sensitivity score of the respondents who selected it. The average sensitivity scores are given for those respondents who did not state that there are no pleasant aspects or that there are no unpleasant aspects. The aspects are ordered according to the average sensitivity. The vertical line represents the overall average noise sensitivity. The figure indicates that noise sensitivity plays a role in the evaluation of many (environmental) aspects. Noise sensitive persons select positive aspects less often as a pleasant aspect of their surroundings. This suggests a general negative attitude. However, the influence of noise sensitivity is not equal for all aspects. It is strongest for the aspects (not) quiet, (no) environmental pollution, and (un) safe living environment. This differentiation, with the strongest effect of noise sensitivity on pure environmental aspects is not consistent with the hypothesis that noise sensitivity is an aspect of a general negative affectivity.

X. CONCLUSION AND DISCUSSION

Noise sensitivity has at most a very weak (positive) relationship with noise exposure, which cannot explain the strong influence of noise sensitivity on effects such as noise annoyance. Weak positive correlations where found in 7 of the 29 currently analyzed datasets. Noise sensitivity changes the influence of noise exposure on noise annoyance, and does not (only) have an additive effect, i.e., it affects the rate at which annoyance increases when the noise exposure gets higher. It also alters reactions other than noise annoyance, such as self-reported sleep disturbance attributed to noise, as well as reactions to other environmental conditions, such as odor. Reactions influenced by noise sensitivity have a strong affective component. If there is no strong affective component (e.g., pure perception), there appears to be little influence of noise sensitivity. Noise sensitivity is more specific than a general critical tendency. It appears that noise sensitivity has relatively little influence on evaluations of nonenvironmental conditions.

The above results suggest that noise sensitive subjects have a predisposition to discriminate environmental conditions and evaluate them. This predisposition is weaker or lacking in persons who are not sensitive to noise. Therefore, the significance of understanding noise sensitivity is not restricted to the field of environmental noise. Hopefully, future work will lead to an integral model of the influence of noise sensitivity on reactions to environmental conditions.

ACKNOWLEDGMENTS

This article has benefited from insightful comments that Jim Fields provided in reviews of drafts of this article. We also thank the anonymous reviewers for their helpful comments.

Belojević, G., and Jakovljević, B. (**1997**). ''Subjective reactions to traffic noise with regard to some personality traits,'' Environ. Int. **23**, 221–226.

Belojević, G., Jakovljević, B., and Santric-Milicević, M. (**1997**). ''Influence of personality on residential behavior in a noise urban area,'' Internoise 97, Budapest, Hungary, 25–27 August, pp. 1249–1251.

Broadbent, D. E. (**1972**). ''Individual differences in annoyance by noise,'' Sound **6**, 56–61.

Dornic, S., and Laaksonen, T. (**1989**). ''Continuous noise, intermittent noise, and annoyance,'' Percept. Mot. Skills **68**, 11–8.

Ellermeier, W., and Zimmer, K. (**1997**). ''Individual differences in susceptibility to the irrelevant speech effect,'' J. Acoust. Soc. Am. **102**, 2191–2199.

Ellermeier, W., Eigenstetter, M., and Zimmer, K. (**2001**). ''Psychoacoustic correlates of individual noise sensitivity,'' J. Acoust. Soc. Am. **109**, 1464–1473.

EU/DG Environment. (**2002**). ''Directive of the European Parliament and of the Council relating to the assessment and management of environmental noise,'' Brussels: EU/DG Environment.

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.

Fields, J. M. (**1993**). ''Effect of personal and situational variables on noise annoyance in residential areas,'' J. Acoust. Soc. Am. **93**, 2753–2763.

Fields, J. M. (**2001**). ''An updated catalog of 521 social surveys of residents' reactions to environmental noise (1943–2000),'' NASA/CR-2001-211257, Hampton, VA, National Aeronautics and Space Administration (http://techreports.larc.nasa.gov/ltrs/PDF/2001/cr/NASA-2001-cr211257.pdf)

ISO (**1987**). ''Acoustics—Description and Measurement of Environmental Noise,'' ISO 1996-2, International Standard Organisation, Geneva.

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.

Job, R. F. S. (**1999**). ''Noise sensitivity as a factor influencing human reaction to noise,'' Noise Health **3**, 57–68.

Job, R. F. S., Hatfield, J., Carter, N. L., Peploe, P., Taylor, R., and Morrell, S. (**1999**). ''Reaction to combined noise sources: the roles of general and specific noise sensitivities,'' Internoise 99. The 1999 International Congress on Noise Control Engineering (Fort Lauderdale, FL), Vol. 2.

Kryter, K. D. (**1982a**). ''Community annoyance from aircraft and ground vehicle noise,'' J. Acoust. Soc. Am. **72**, 1212–1242.

Kryter, K. D. (**1982b**). ''Community annoyance from aircraft and ground vehicle noise'' (Response of K. D. Kryter to modified comments by T. J. Schultz on K. D. Kryter's paper). J. Acoust. Soc. Am. **73**, 1066–1068.

Langdon, F. J. (**1976**). ''Noise nuisance caused by road traffic in residential areas: part III,'' J. Sound Vib. **49**, 241–256.

Lercher, P. (**1995**). ''Distinguishing effects modifiers of psychological morbidity associated with traffic noise exposure,'' Epidemiology **S28**, P085.

McKennel, A. C. (**1963**). ''Aircraft noise annoyance around London (Heathrow) Airport'' (Her Majesty's Stationary Office, London).

Meijer, H., Knipschild, P., and Salle, H. (**1985**). ''Road traffic noise annoyance in Amsterdam,'' Int. Arch. Occup. Environ. Health **56**, 285–297.

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., Vos, H., and de Jong, R. G. (**2000**). ''Community reaction to aircraft noise: time-of-day penalty and tradeoff between levels of overflights,'' J. Acoust. Soc. Am. **107**, 3245–3253.

Moreire, N. M., and Bryan, M. E. (**1972**). ''Noise annoyance susceptibility,'' J. Sound Vib. **21**, 449–462.

Öhrström, E., Rylander, R., and Björkman, M. (**1988**). ''Effects of night time road traffic noise an overview of laboratory and field studies on noise dose and subjective noise sensitivity,'' J. Sound Vib. **127**, 441–448.

Raw, G. J., and Griffith, I. D. (**1988**). ''Individual differences in response to road traffic noise,'' J. Sound Vib. **121**, 463–471.

Schultz, T. J. (**1978**). ''Synthesis of social surveys on noise annoyance,'' J. Acoust. Soc. Am. **64**, 377–405.

Stansfeld, S. A. (**1992**). ''Noise, noise sensitivity and psychological studies,'' Psychol. Med. Monograph Suppl. 22 (Cambridge U.P., Cambridge).

Stephens, S. D. G. (**1970**). ''Personality and the slope of loudness functions,'' Q. J. Exp. Psychol. **22**, 9–13.

Tabachnick, B. G., and Fidell, L. S. (**1983**). *Using Multivariate Statistics* (Harper & Row, New York).

Thomas, J. R., and Jones, D. M. (**1988**). ''Individual differences in noise annoyance and the uncomfortable loudness level,'' J. Sound Vib. **82**, 289–304.
Waddell, P. A., and Grondwall, D. M. A. (**1984**). ''Sensitivity to light and sound following minor head injury,'' Acta Neurol. Scand. **69**, 270–276.
Watson, D., and Clark, L. E. (**1984**). ''Negative affectivity: The disposition to experience aversive emotional states,'' Psychol. Bull. **96**, 465–490.
Watson, D., and Pennebaker, J. W. (**1989**). ''Health complaints, stress and distress: Exploring the central role of negative affectivity,'' Psychol. Rev. **96**, 234–254.
Weinstein, N. D. (**1978**). ''Individual differences in reactions to noise: a longitudinal study in a college dormitory,'' J. Appl. Physiol. **63**, 458–466.
Winneke, G., and Neuf, M. (**1992**). ''Psychological response to sensory stimulation by environmental stressors: trait or state?'' Appl. Psychol. **41**, 257–267.
Zimmer, K., and Ellermeier, W. (**1997**). ''Eine deutsche Version der Lärmempfindlichkeitsskala von Weinstein [A German version of Weinstein's noise sensitivity scale],'' Z. Lärmbekämpfung **44**, 107–110.
Zimmer, K., and Ellermeier, W. (**1998**). ''Konstruktion und Evaluation eines Fragebogens zur Erfassung der individuellen Lärmemfindlichkeit [Construction and evaluation of a questionnaire for the assessment of individual noise sensitivity],'' Diagnostica **44**(1), 11–20.
Zimmer, K., and Ellermeier, W. (**1999**). ''Psychometric properties of four measures of noise sensitivity: a comparison,'' J. Environ. Psychol. **19**, 295–302.

Measurement of acoustical characteristics of mosques in Saudi Arabia

Adel A. Abdou[a)]
Department of Architectural Engineering, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

(Received 19 November 2001; revised 24 May 2002; accepted 4 November 2002)

The study of mosque acoustics, with regard to acoustical characteristics, sound quality for speech intelligibility, and other applicable acoustic criteria, has been largely neglected. In this study a background as to why mosques are designed as they are and how mosque design is influenced by worship considerations is given. In the study the acoustical characteristics of typically constructed contemporary mosques in Saudi Arabia have been investigated, employing a well-known impulse response. Extensive field measurements were taken in 21 representative mosques of different sizes and architectural features in order to characterize their acoustical quality and to identify the impact of air conditioning, ceiling fans, and sound reinforcement systems on their acoustics. Objective room-acoustic indicators such as reverberation time (RT) and clarity (C_{50}) were measured. Background noise (BN) was assessed with and without the operation of air conditioning and fans. The speech transmission index (STI) was also evaluated with and without the operation of existing sound reinforcement systems. The existence of acoustical deficiencies was confirmed and quantified. The study, in addition to describing mosque acoustics, compares design goals to results obtained in practice and suggests acoustical target values for mosque design. The results show that acoustical quality in the investigated mosques deviates from optimum conditions when unoccupied, but is much better in the occupied condition.
[DOI: 10.1121/1.1531982]

PACS numbers: 43.55.Br, 43.55.Gx, 43.55.Jz, 43.55.Mc [MK]

I. INTRODUCTION

Mosques are places of worship used for prayer, public speaking, preaching, lecturing, and *Qur'an* recitations. All activities performed in mosques are related to speech audibility and intelligibility. Therefore, the design of their acoustical features requires careful consideration if good listening conditions are to be achieved. Although mosques are uniquely important buildings in every Muslim community, in general, their acoustical quality evaluation, problems (if any) and possible remedies have not received adequate attention in the literature. Hammad[1] in an early study evaluated speech intelligibility via rapid speech transmission index (RASTI) measurements in mosques in Amman, Jordan. He concluded that, in general, the acoustical characteristics of mosques had been largely neglected. In 1991, the acoustical problems of a huge mosque built in Amman were investigated.[2] The authors recommended that acoustical properties of mosques should be considered at the early stages of design. In 1995, a study[3] established the relative influence of active environmental control systems on the acoustical performance of a typical mosque in the Gulf region. Recently, Abdou[4] presented and discussed the potential of utilizing room-acoustics simulation in the early stages of mosque design, where decisions are made to establish the mosque geometry, surface materials, and speech reinforcement system (SRS) distribution.

Acoustic evaluations of other religious buildings[5–9] have been extensively reported. For example, the varying of the acoustics of a large cathedral for satisfactory speech intelligibility, by the use of carefully designed, installed, maintained, and operated sound amplification system has been demonstrated and discussed.[5] In addition to assessing acoustical quality by using pressure-based room acoustics indicators, visualizing the directional characteristics of sound fields at the listener position is also possible. Three-dimensional transient sound intensity impulse responses have been utilized to assess the effectiveness of a sound system in a large reverberant church.[7] Subjective and objective acoustical field measurements have been conducted in a survey[9] of 36 Roman Catholic churches in Portugal. The idea was to evaluate and predict the acoustical quality of these churches. Correlation analyses and statistical modeling identified relationships between some room-acoustic indicators and speech intelligibility in this particular style of church. Recently clarity and definition acoustic indices in Gothic churches were measured and compared with expected results derived from a semi-empirical analytical model.[10] In the literature, objective and subjective evaluation of halls used for other functions such as concert halls, opera houses and classrooms have been widely reported. However, developments employing impulse response techniques[11,12] for evaluating the acoustical quality in different types of enclosures have not yet been applied to mosques, at least not in widely known publications. In addition not many readers are aware of mosque design, its worship considerations, acoustical properties, and requirements.

The objectives of the current work were the following:

[a)]Assistant Professor of Architectural Engineering. Electronic mail: adel@kfupm.edu.sa

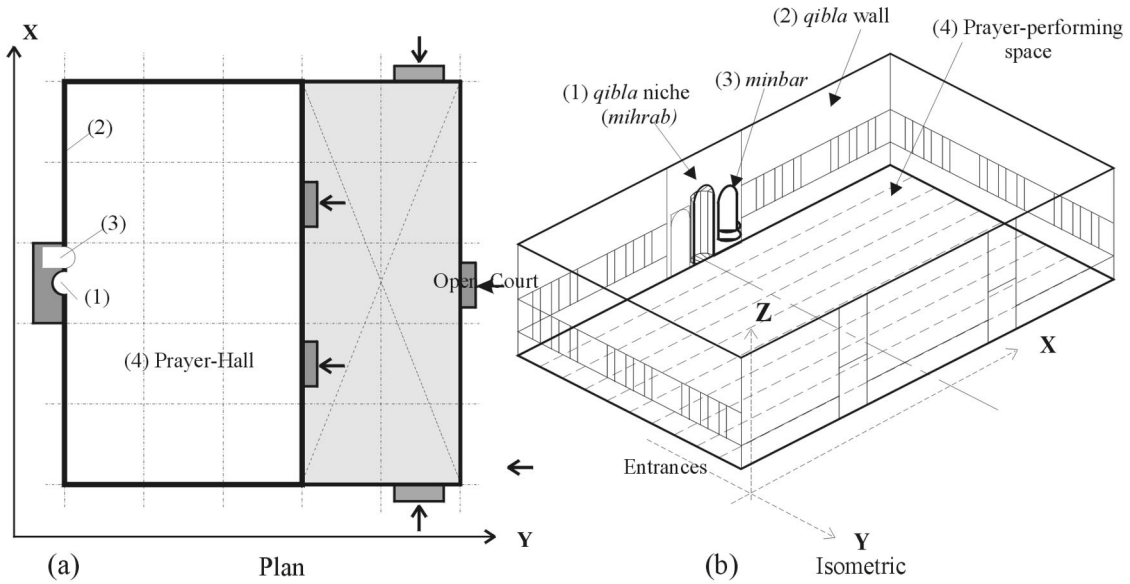


FIG. 1. The basic design elements of a simple mosque (a) plan, and (b) isometric.

(1) To give a background about the mosque's basic design elements as influenced by worship considerations and mosque classifications;
(2) to characterize the acoustical quality of typical contemporary mosques built in Saudi Arabia and, subsequently, to objectively confirm (or otherwise) the existence of acoustical problems with respect to speech intelligibility and compare design goals with results obtained in practice;
(3) to identify the impact of active environmental control systems such as air-conditioning systems (A/C) and mechanical ventilation incorporated into mosque designs on the acoustical quality; and
(4) to investigate the overall effectiveness of the SRS most commonly operated in mosques for enhancing speech intelligibility.

II. COMMON CHARACTERISTICS OF CONTEMPORARY MOSQUES

A. Basic design elements of a mosque

Historically, the first mosque built in *Al-Madinah Al-Monawarah* city, *Saudi Arabia* formed the model for subsequent mosques throughout the Islamic world[13] in its combination of basic elements. It was a simple rectangular, walled enclosure with a roofed prayer hall. The long side of the rectangle is oriented toward the direction of the holy mosque in *Makkah* city. This wall is usually described as the *qibla* wall. The wall contains a recess in its center in the form of a wall niche called the *mihrab*. This wall also includes the *minbar* which is commonly an elevated floor, to the right of the *mihrab*, from which the *Imam* preaches or delivers the *Friday* sermon, the *khutba*. These basic elements are the essentials of mosque design in Saudi Arabia, as they are elsewhere in the Islamic world. Figure 1 illustrates the plan and isometric of a simple, typical mosque design. The basic design elements are emphasized. Since the construction of the first mosque, the functions of every mosque have remained unchanged. However, the mosque architectural form, space, construction system, and building materials have evolved and developed to a significant and variable extent in different parts of the Islamic world, influenced by many factors mentioned elsewhere.[13,14]

B. Worship modes in a mosque

The mosque design is mainly influenced by worship considerations. Worship in a mosque consists of two major modes. The first mode, namely the *prayer mode*, involves performing prayers either individually or in a group, as religiously prescribed. Group prayer must be performed with worshippers standing, bowing, prostrating, or sitting behind the *Imam*, on the same floor level, aligned in rows parallel to the *qibla* wall with distances around 1.2 m apart. The second mode is the *preaching mode*, where worshippers are directly seated on the floor in random rows listening to the *Imam* preaching or delivering the *khutba* while standing on the elevated *minbar* floor. The *minbar* floor height varies from one mosque to the other but usually is in the range of one to three meters above the mosque floor. Figure 2 shows the worshippers' different postures and their orientation in relation to the *Imam* while performing the two different religious activities in the mosque. The congregational capacity of the mosque is usually determined by the floor area divided by the area required for a worshipper to perform the prayer, i.e., approximately $0.80\times1.2=0.96$ m^2.

C. Mosque classifications

While, in general, *traditional* mosques can be classified according to their architectural form and configuration, *contemporary* mosques may be broadly classified according to their size and location in relation to the community.[14,15] Large mosques are located in large cities as public landmarks. They are usually built by the government expressing the state's commitment to Islam. They are generally grand in size and of large congregational capacity. Community

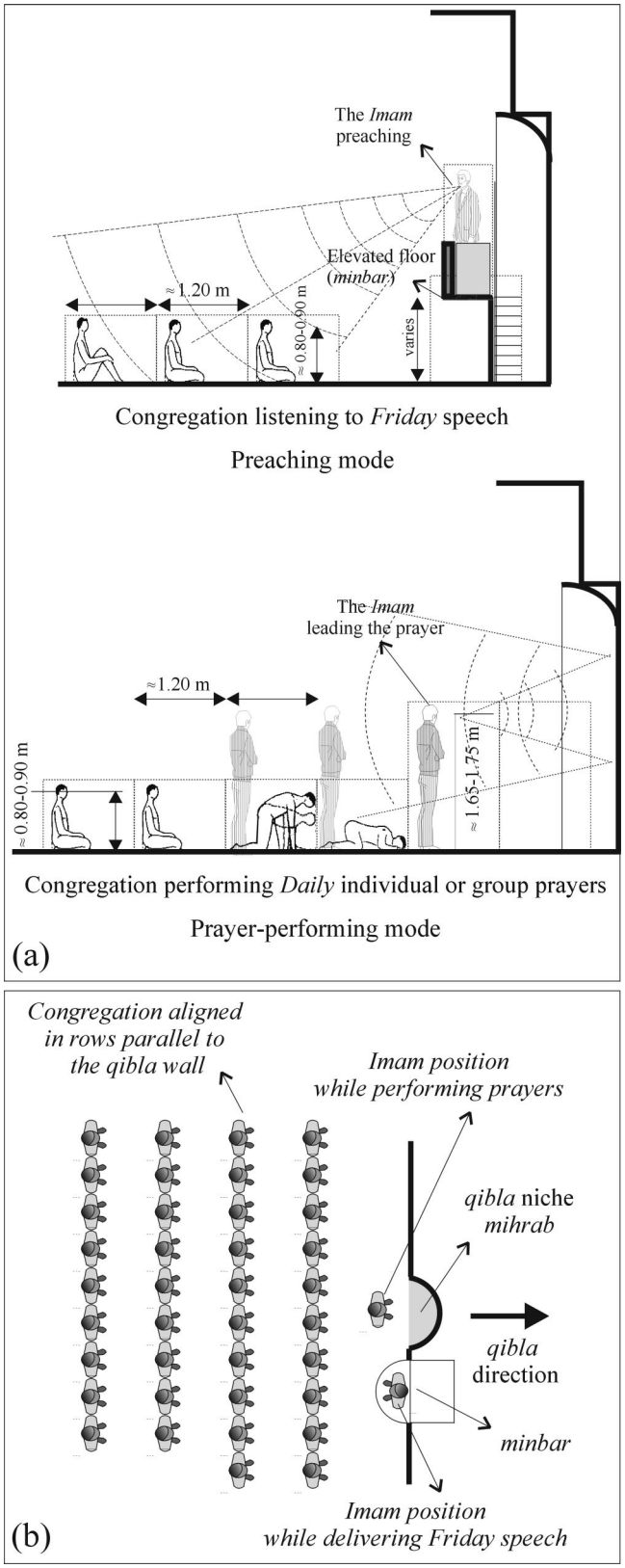


FIG. 2. The worshippers' different postures and their orientation in relation to the *Imam* in the two religious modes, (a) sections showing congregations listening to the *Friday* speech (i.e., the preaching mode) and performing *Daily* individual or group prayer (i.e., the prayer-performing mode), and (b) a top-view plan showing the source–receiver path (i.e., *Imam* worshippers) in group prayer performing.

mosques (i.e., *Jamma* mosques, where *Friday* prayer, preceded by *Friday* speech, can be performed) are distributed throughout urban and rural communities and may house additional functions (e.g., a library, meeting rooms, clinic, etc.), in addition to the prime function of a space for performing prayers. They are usually utilized for both *daily* prayers as well as the *Friday* prayer and occasionally are supplemented with a separate annex on the same floor level or in a mezzanine floor for female worshippers. Small local mosques are located in small neighborhoods, and are of modest dimensions and congregational capacity. The planning and design guidelines for the above three types of mosque are available in Ref. 15.

D. Mosque prototypes in Saudi Arabia: Common features

In Saudi Arabia, many prototypes of mosque design exist.[16] Mosques are built in various sizes ranging from small and medium to large types. They are usually typical in layout, shape, construction system, and building material, but with different types of air conditioning and electroacoustic sound systems. From a field survey of 90 mosques, it was observed from site visits, design drawings, and ''as built'' sketches that mosques are fairly similar with respect to their construction systems. They are commonly constructed of reinforced concrete skeletal structures with flat roofs. The flat roof is commonly supported on columns that are arranged on a regular grid (i.e., structural unit). A dome is sometimes constructed spanning the central part of the roof to eliminate intermediate columns. The shape represented by the aspect ratio (i.e., the mosque length over width), and the floor area of each mosque type is mainly controlled by the size and proportion of the structural unit dimensions as well as the total number of units (e.g., 5×3 units). Interior materials of these typical contemporary mosques vary. Walls are mostly finished with reflecting materials such as painted plaster. They usually contain a wainscot, around 1.0 m high, made of marble tiles. The floor area is always covered with heavy carpet. Hard, painted concrete ceilings with simple to elaborate decorations are commonly used. Due to the harsh climatic conditions in most of Saudi Arabia's regions, air-conditioners are virtually a necessity. Therefore, almost all types of mosques are equipped with either a central, or a split unit air-conditioning system or window-type unit, in concert with ceiling fans for air circulation. Electroacoustic SRS have also been implemented in mosques of all sizes to enhance the listening conditions in the mosque space, particularly after the introduction of the air-conditioning systems and the anticipated subsequent increase of ambient noise in the mosque.

III. ASSESSMENT OF ACOUSTICAL QUALITY IN MOSQUES

Nowadays, numerous subjective attributes of the listening experience in enclosures can be described by the many available contemporary room-acoustic indicators. A comprehensive listing of these contemporary indicators, definitions, corresponding subjective attributes, and suggested tolerance

range values can be found in Ref. 17. In mosques, the major acoustical concern is verbal communication. All activities in the mosque are dependent on speech audibility and intelligibility. Speech intelligibility (SI) is the percentage of speech material that is correctly identified by the average listener. The intelligibility of speech in rooms is related to both the speech signal-to-noise ratio and to the acoustic characteristics of the enclosure. That is, it can be influenced by the speech sound level, ambient background noise (BN), and the reverberation time (RT) of the enclosure. Reverberation affects SI by affecting the early-to-late arriving sound energies. SI is directly related to the early-to-late energy fraction, namely sound definition. In summary, in order to measure or predict speech intelligibility, various objective-based measures can be used. Definition (D_{50}) which is related to sound clarity (C_{50}), useful-to-detrimental sound ratios (e.g., SNR_{95},[18] U_{50}, and U_{80}[19,20]), Speech transmission index (STI),[21] rapid speech transmission index (RASTI),[22,23] and the articulation loss of consonants ($\%AL_{cons}$)[24] are all indicators of speech intelligibility with varying degrees of accuracy. Many studies have investigated and compared measures of SI in rooms,[25–27] particularly in classrooms.[28–30] In this study RT, BN, C_{50}, and STI were measured to characterize the acoustical conditions in mosques and assess SI.

IV. METHODOLOGY FOR OBJECTIVE FIELD MEASUREMENTS

The first step in this investigation was to select sample mosques representative of the majority of existing mosque prototypes in Saudi Arabia. Ninety mosques were field surveyed to assess their actual design and operation status. An audit form for collecting relevant and essential data such as the ''as-built'' physical information of mosques during site visits was developed. The audit form addressed major aspects such as the following: general information about the mosque, including a given reference number and location; physical data, including the number of structural units and dimensions; and use whether for *daily* or *Friday* prayers, or for both, and the existence of a separate hall or area for women to perform prayers. Due to the importance of identifying possible interior sources of noise as well as quantifying ambient noise in the mosque, the type and unit distribution of the A/C system were also considered. The mosque's existing SRS in terms of number of used loudspeakers, type, height, and spatial arrangement was documented. A criterion was then established to select representative sample mosques. The dominance of mosque shape, size (type repetition), and other factors contributed to the final selection in addition to mosque accessibility and proximity for the ease of equipment movement. Out of the total of 90 mosques, 21 (i.e., a approximately 23%) were selected for acoustical field measurements. Subsequently, pilot experimental measurements in one of the selected mosques were implemented in order to develop detailed procedures for the measurements. After refinement of the measurement setup, systematic acoustical measurements in all the selected sample mosques were then carried out and results analyzed.

A. Grouping of the sample mosques

Considering the mosque volume as an important parameter influencing acoustical characteristics, the sample mosques were grouped in five categories. Table I presents data summarizing the main physical characteristics of the chosen 21 mosques. It includes information such as number of structural units, mosque's length, width, height, area, volume, and congregational capacity, all sorted in ascending order with respect to the mosque volume. Group (A) includes very small local mosques with volume $<1000\ m^3$; Group (B) includes small mosques with volume $\geqslant 1000\ m^3$ and $<1500\ m^3$; Group (C) includes medium mosques $\geqslant 1500\ m^3$ $<2000\ m^3$; Group (D) contains large mosques with volume $\geqslant 2000\ m^3$ and $<3000\ m^3$; and Group (E) contains very large mosques with volume $\geqslant 3000\ m^3$. The architectural plans of the selected mosques grouped into five categories as identified above and indicated in Table I are shown later in part (a) of Fig. 11. The selected mosques varied from very small local mosques with an average volume of around $630\ m^3$ and a capacity of as few as 140 worshippers to large mosques with volumes over $3000\ m^3$ and a capacity of over 800 worshippers. Mosque shape is represented by the aspect ratio (AR), i.e., length, *L* of *qibla* wall over mosque width, *W*, capacity, and volume per worshipper is also indicated. The mosque shapes varied from a square (AR=1) to an elongated rectangular shape (AR=2.7) with an average AR of 1.7. All had rectangular cross-sectional shapes of various aspect ratios. Group (F) includes a single huge (i.e., landmark) mosque with a volume greater than $10000\ m^3$ and an occupancy of over 2800 worshippers, used for *Friday* prayer only. The minimum, average, and maximum value range and standard deviation (STD), excluding the mosque referenced DM43 in Group (F), of the physical characteristics of the sample mosques, are shown in Table I.

B. Measurement system and procedures

In order to characterize and evaluate the acoustical characteristics of the selected mosques the maximum length sequence system analyzer (MLSSA)[31] was utilized. It is a PC-based acoustic measuring system and analyzer for the measurement and evaluation of room acoustics. MLSSA employs a maximum-length sequence (MLS) for the excitation signal as a preferred alternative to the conventional white noise stimulus. The MLS signal technique measures the impulse response—the most fundamental descriptor of any linear system—from which a wide range of important acoustic indicators can be determined through computer-aided post-processing. MLSSA was used for the measurement, in combination with an ''omnidirectional'' sound source of dodecahedral configuration. The sound source was located in the center of the *qibla* niche 1.0 m away from the *qibla* wall. The height of the excitation sound source was maintained at 1.55 m from the floor to represent a person talking in a standing position. This source was used to determine RT_{30}, EDT, and C_{50}. Since SI is affected by the directivity of the sound source, a small test loudspeaker was used as the sound source radiating with sound directivity approximating that of a human speaker for STI and $\%AL_{cons}$ measurements to im-

TABLE I. Summary of main physical characteristics of the selected sample mosques sorted in ascending order with respect to mosque's volume. Range, average, and standard deviations are also indicated.

No.	Mosque reference	USE[a]	# of structural units	On-site measurements: Dimensions (geometry) L	W	H	Calculated parameters: Aspect ratio L/W	Area (A) m^2	Capacity (C)[b] Person (P)	Volume (V) m^3	Groups/volume	V/C m^3/P	Mean m^3/P
1	TH16	D	3×3	11.55	11.55	3.90	1.0	133	139	520	Group (A)	3.7	4.2
2	TH32	D	3×2	13.57	9.65	4.49	1.4	131	136	588	(<=1000 m^3)	4.3	
3	DM242	D	3×2	15.00	10.00	4.50	1.5	150	156	675	mean V=630[c]	4.3	
4	TH27	D	3×3	13.49	12.04	4.60	1.1	162	169	747		4.4	
5	TH48	D	5×3	24.65	14.66	3.33	1.7	361	376	1203	Group (B)	3.2	4.3
6	DM16	FD	1/2,3,1/2×3	19.50	14.70	4.25	1.3	287	299	1218	(>1000<=1500 m^3)	4.1	
7	DM260	D	3×3	18.00	15.00	4.61	1.2	270	281	1245	mean V=1290	4.4	
8	KH45	D	5×2	24.85	9.80	5.63	2.5	244	254	1371		5.4	
9	KH17	FD	5×4	19.73	15.75	4.52	1.3	311	324	1405		4.3	
10	KH03	FD	7×4	24.25	13.70	4.70	1.8	332	346	1561	Group (C)	4.5	4.2
11	TH42	D	3×3	24.20	18.35	4.00	1.3	444	463	1776	(>1500<=2000 m^3)	3.8	
12	KH12	D	9×3	35.14	13.22	4.20	2.7	465	484	1951	mean V=1820	4.0	
13	DH14	FD	1/2,5,1/2×3	29.56	15.33	4.40	1.9	453	472	1994		4.2	
14	DM125	D	5×3	25.00	15.00	5.40	1.7	375	391	2025	Group (D)	5.2	4.7
15	KH59	FD	5×3	24.80	14.85	5.65	1.7	368	384	2081	(>2000<=3000 m^3)	5.4	
16	TH06	FD	7×3	37.45	15.88	3.70	2.4	595	619	2200	mean V=2200	3.6	
17	DM06	FD	7×3	32.10	15.80	4.94	2.0	507	528	2505		4.7	
18	DH03	FD	9×4	43.43	19.43	6.05	2.2	844	879	5105	Group (E)	5.8	5.0
19	TH13	FD	7×6	42.25	29.68	4.80	1.4	1254	1306	6286	(>3000<=10 000 m^3)	4.8	
20	TH01	FD	9×7	44.85	34.70	4.52	1.3	1556	1621	7034	mean V=6140	4.3	
21	DM43	F	9×9	52.00	52.00	8.65	1.0	2704	2817	23 390	Group (F) (>10 000 m^3)	8.3	8.3
Minimum[d]				11.6	9.7	3.3	1.0	131	136	520		3.2	4.2
Average[d]				26.2	16.0	4.6	1.7	462	481	2175		4.4	4.5
Maximum[d]				44.9	34.7	6.1	2.7	1556	1621	7034		5.8	5.0
STD[d]				10.0	6.0	0.7	0.5	359	374	1780		0.6	0.3

[a]D=*Daily* prayers only, FD=*Friday* and *Daily* prayers.
[b]Capacity is calculated by considering a required area of (0.80×1.20=0.96 m^2) for each worshipper.
[c]Average volume for each group to the nearest 10 m^3.
[d]Excluding DM43, Group (F).

prove results. The small loudspeaker was located at the typical *Imam* praying position.

Based on the shape and floor area of each sample mosque, an adequate number of listener positions was selected for measurement in order to achieve a proper coverage of the mosque floor area. The guidelines set by the ISO 3382 (1997)[32] were adhered to. Impulse responses were then measured sequentially in each of the chosen listener locations. Measurements were acquired using a calibrated 1/2 in. condenser microphone mounted on an adjustable microphone holder fixed to a tripod, maintained 1.65 m above the floor representing the location of a listener's ear in a standing position. The measurement was also repeated in the same listener positions but with the microphone height lowered to a height of 0.85 m above the floor representing the location of the ear of a listener in a floor-seated position. Measurements were carried out with the mosque unoccupied, and with both the ceiling fans and the A/C system set to the ''OFF'' condition. The average of measurements taken on the two different heights at all listener locations in the mosque would represent worshippers at the two worship modes of preaching and prayer performing and therefore would characterize the spatial value of the acoustical indicator.

It was necessary to measure the mosque BN and subsequently determine the noise criterion (NC) rating. The octave-band sound pressure level (SPL) of ambient BN was measured at each selected measurement location using the same calibrated 1/2 in. condenser microphone maintained at 1.65 m above the floor. The same was conducted at the measurement point located in the center of the mosque floor area with all ceiling fans operating. A wind screen was used to reduce the effect of airflow due to the operation of the ceiling fans. The measurement of BN was repeated with only the A/C system set to ''ON'' and then with the A/C system and fans operating concurrently The idea was to quantify the effect of these active environmental control systems on the magnitude and spectral content of BN.

With the use of the small test loudspeaker, STI and RASTI were also measured from the acquisition of long-duration impulse responses (>1 s). The A-weighted SPL of the generated MLS signal at a location 1.0 m directly in front of the test loudspeaker was adjusted to achieve an A-weighted sound of 67–68 dB (A). A 65535-point impulse response was then acquired, and analyzed for STI calculations. To determine the overall STI value, the original

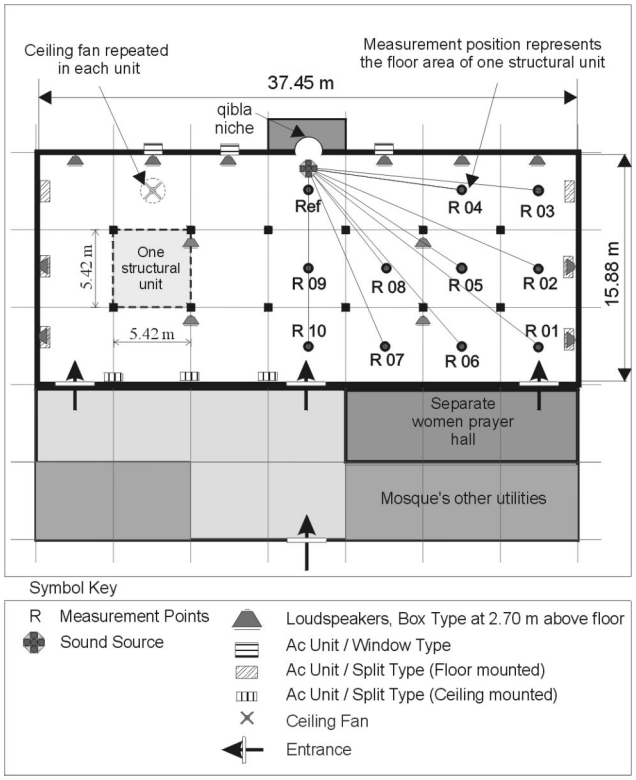


FIG. 3. Plan view of the *TH06* mosque showing a geometric configuration of the prayer hall, A/C unit distribution, and the mosque's SRS wall-mounted loudspeaker distribution. The measurement positions and sound source location are also shown.

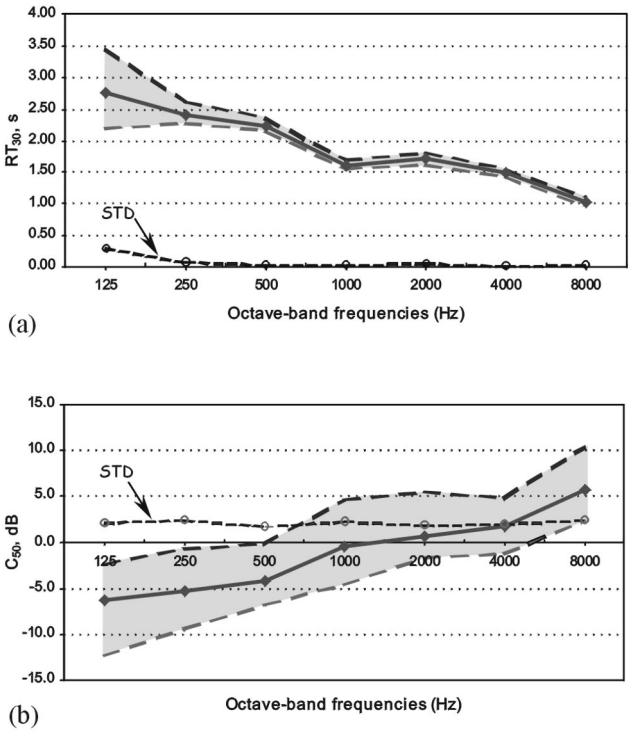


FIG. 4. The spatial average (♦) of (a) RT_{30} and (b) C_{50} measured in the *TH06* mosque. The shaded area represents the measured value range (i.e., spatial minimum and maximum).

weighting factors given by Steeneken and Houtgast[33] were used. In addition, the modified weighting factors given by French and Steinberg were also utilized, resulting in a modified version of STI denoted $STI_{modified}$.[31,34]

SRS commonly used in a mosque consists of three or more microphones for picking up the *Imam's* voice, preamplifiers, a power amplifier, and several loudspeakers that reradiate the amplified speech sounds. Additional components like mixers and equalizers are added to the basic SRS components in very few mosques. Many factors affect the performance of SRS such as the acoustical power, quality of the system components, loudspeakers directivity and frequency response, and reverberation time of the room in which the system is operated. The purpose was to determine the overall effectiveness of installed speech reinforcement systems as set and operated in the chosen sample mosques. Measurements were conducted at each of the chosen listener positions without operating the mosque's SRS. The SRS was then put ''ON'' and the MLS signal was sent via the amplifier (when accessible) or via the microphone of the mosque's SRS. The volume controls and component settings of the system were kept as usually set without any alteration. The impulse responses were captured at a height of 0.85 m representing worshippers seated on the floor listening to the *Imam*.

C. Sample field measurements and analysis: Unoccupied mosque

Pilot field measurements were conducted in one mosque referenced TH06, Group (D) (see Table I) to test the measurement setup for appropriateness, to identify potential difficulties, and to assess on-site data analysis. Figure 3 depicts the mosque's geometry, physical characteristics, and measurement positions performed in the mosque. The mosque's shape being symmetric, ten measurement points were examined on one side of the floor area. These are denoted R01 to R10. Each measurement location was considered to represent the floor area of one structural unit as defined on the plan view of the mosque. Measurement positions were selected to represent listeners in the front, center, and back areas of the floor. It was ensured, while selecting the measurement locations, that the distance between any two selected listeners' locations was more than two meters, and the distance from the microphone position to the nearest reflecting surface was more than one meter, (i.e., approximately equal to half the wavelength of the lowest octave-band frequency of interest, 125 Hz). Measurement positions were also considered with regard to existing A/C units and the location of the loudspeakers of the mosque's SRS.

Figure 4(a) depicts the spatial minimum, average, and maximum values of RT_{30} i.e., the value range of the mosque's RT_{30} spectrum in octave-band frequencies from 125 Hz to 8 kHz resulting from the processing of the 20 impulse responses captured at heights of 0.85 and 1.65 m above the floor. As can be seen, the spatial-averaged RT at mid-frequencies (500–1000 Hz), $RT_{30\,m}$ is 1.90 s with a negligible STD. From a knowledge of the mosque volume (2200 m^3), the optimum $RT_{30\,m}$ value for speech is approximately 0.90 s. Therefore the mosque can be said to have a high RT when unoccupied. In the literature, there is no specific data of the absorption coefficients of worshippers or audience standing or seated directly on a carpeted floor in rows 1.2 m

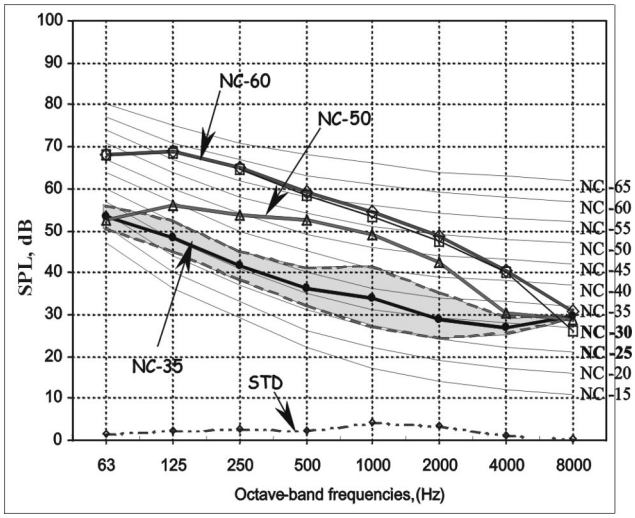


FIG. 5. Spatial average of BN spectra (●) measured in the *TH06* mosque at different operating conditions A/C "ON" (△) Fans "ON" (—) and A/C and Fans "ON" (◇). The shaded area represents the measured value range (i.e., spatial minimum and maximum).

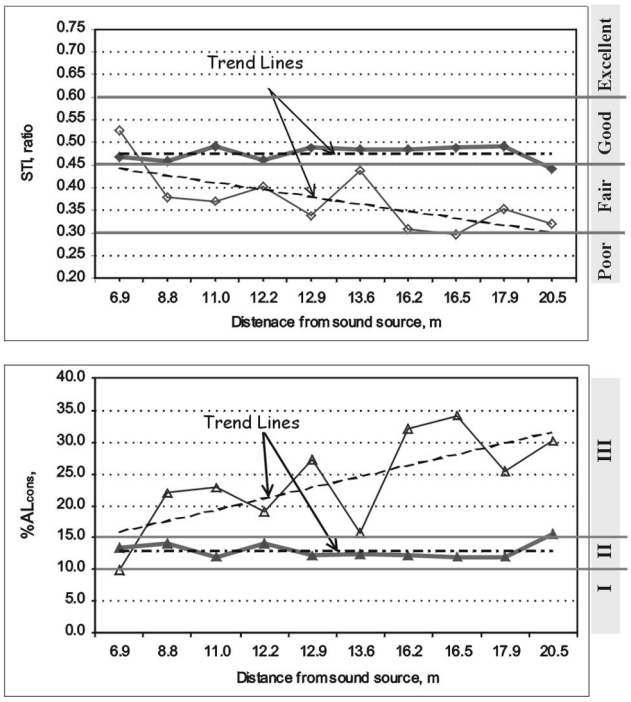


FIG. 6. (a) A comparison of STI measured at all positions with (♦) and without (◇) the operation of the SRS, and (b) a comparison of $\%AL_{cons}$ values versus distance from the sound source with (▲) and without (△) the operation of the mosque's SRS. SI ratings are shown on the right *y* axis. I $=(\%AL_{cons}<10\%)$ very good, II$=(\%AL_{cons}>10\%<15\%)$ good, and III $=(\%AL_{cons}>15\%)$ insufficient.

apart. Assuming the absorption area of worshippers when standing in parallel rows 1.2 m apart is around 0.4 metric sabines per person at mid-frequencies (500–1000),[35] then the mosque's RT would be approximately 1.32, 1.01, and 0.82 s with one-third of the mosque occupied, two-thirds and fully occupied, respectively. The measured RT would therefore be optimal with the mosque occupied up to two-thirds of its congregational capacity. The spatial-averaged value of the mid-frequency C_{50} as shown in Fig. 4(b) was −2.4 dB, varying from −5.1 to 1.4 dB with a STD of approximately 2.0 dB. A C_{50} value higher than 1.0 dB or more is required for satisfactory speech intelligibility. This value corresponds to a sound definition, D_{50}, of about 0.56.

Both the RT and the ambient BN affecting speech intelligibility were measured, in standard octave-band frequencies, at all selected measurement locations in different operating conditions. Both the overall A-weighted and Linear SPL were determined along with the NC rating. The spatial minimum, average, and maximum values of all measurements were then calculated. These spectra ranges are illustrated in Fig. 5. Operating either the fans or the A/C only or both proved that the BN increased from NC35 to NC60. The spatial-averaged spectrum of the BN in a mosque can be considered quite high compared to the recommended value range of NC25–NC30 for rooms intended for speech such as lecture halls and classrooms. It can be expected that the sound quality will further deteriorate when the ceiling fans or the A/C units or both are operated. The causes of the high NC rating of ambient noise can be attributed to intruding exterior noise due to the low transmission loss of the exterior wall system, particularly windows and glazing types commonly installed in mosques, in addition to interior noises such as buzzing and humming resulting from defective lighting units and accessories.

Figure 6 shows the STI values along with the $\%AL_{cons}$ and SI ratings measured at all selected measurement positions with and without the operation of the mosque's SRS. Measurement positions are shown in ascending order with respect to distance from the sound source. The SI average rating was be poor without the operation of the mosque SRS. Figure 6(a) shows a comparison of STI measured values versus the distance from the sound source with and without the SRS being operated. The STI values improved when the SRS was used, shifting the SI rating to a higher range in almost all measurement locations. $STI_{modified}$ values showed the same trend, however, this modified parameter was found not to contribute to a significant difference or a different evaluation of SI rating for the case under study. Since $STI_{modified}$ is not also in general use as STI, only STI results will be reported in this study. The $\%AL_{cons}$ also shows the same trend of improvement at all locations except measurement location R09. Usually the SRS is installed to achieve an adequate sound level in areas of the mosque remote from the *Imam* location. The $\%AL_{cons}$ values versus distance from the sound source with and without using the SRS are illustrated in Fig. 6(b). It further confirms the efficiency of the SRS of reducing the $\%AL_{cons}$, particularly in the more remote locations. It can be concluded from the above measurement results that the mosque's acoustical characteristics and quality are not satisfactory for speech intelligibility when used without the operation of the SRS. The SRS was found to be quite efficient in improving speech intelligibility from *poor* to *fair* and reducing the $\%AL_{cons}$, particularly in the more remote locations. However, when operating either the fans or the A/C system, the ambient noise in the mosque significantly increases to unacceptable levels and SI can be expected to degrade.

TABLE II. Spatial-average spectra of RT_{30} and C_{50} for each mosque group. The Mid-frequency average is shown in the last two columns. Note that Group (F) is excluded from calculating the value range (minimum, average, and maximum).

	Octave-band frequencies (Hz)							Average	
Groups	125	250	500	1000	2000	4000	8000	500–1000	500–2000
Group (A)	2.79	2.54	2.28	1.64	1.18	1.10	0.77	2.0	1.7
Group (B)	2.26	1.71	1.37	1.25	1.19	1.10	0.86	1.3	1.3
Group (C)	2.26	1.36	0.97	1.32	1.42	1.31	1.06	1.1	1.2
Group (D)	2.66	2.13	1.83	1.49	1.36	1.20	0.89	1.7	1.6
Group (E)	2.89	2.70	2.44	1.95	1.69	1.50	1.08	2.1	2.0
Group (F)	2.91	2.74	2.85	2.47	1.91	1.56	1.01	2.6	2.4
Minimum	2.3	1.4	1.0	1.2	1.2	1.1	0.8	1.1	1.2
Average	2.6	2.1	**1.8**	**1.5**	1.4	1.2	0.9	**1.6**	1.6
Maximum	2.9	2.7	2.4	1.9	1.7	1.5	1.1	2.1	2.0
STD	0.3	0.5	0.5	0.2	0.2	0.2	0.1	0.4	0.3
Group (A)	−6.0	−5.1	−4.1	−1.2	0.4	2.0	5.4	−2.7	−1.7
Group (B)	−3.3	−1.9	−0.2	0.6	0.9	2.3	4.8	0.2	0.4
Group (C)	−2.9	−0.3	1.0	0.5	0.7	1.7	3.6	0.8	0.7
Group (D)	−4.7	−4.1	−3.0	−0.2	1.0	2.2	5.5	−1.6	−0.7
Group (E)	−6.6	−6.6	−5.3	−2.5	−0.9	0.4	4.2	−3.8	−2.9
Group (F)	−8.1	−8.8	−9.2	−3.8	−1.3	0.0	4.4	−6.5	−4.7
Minimum	−6.6	−6.6	−5.3	−2.5	−0.9	0.4	3.6	−3.8	−2.9
Mean	−4.7	−3.6	**−2.3**	**−0.6**	0.4	1.7	4.7	**−1.4**	−0.8
Maximum	−2.9	−0.3	1.0	0.6	1.0	2.3	5.5	0.8	0.7
STD	1.5	2.3	2.4	1.2	0.7	0.7	0.7	1.7	1.3

V. RESULTS COMPARED TO DESIGN GOALS FOR ALL SAMPLE MOSQUES

The type of measurement and analysis described above for one single mosque was systemically performed for all the sample mosques. The purpose was to document the acoustical characteristics of mosques and compare design goals with obtained results in practice. This eventually would indicate the overall acoustical quality of the different prototypes of the selected mosques and how significant they deviate from optimal.

A. Acoustical characteristics of mosques

To summarize the measurement results of all sample mosques in a comparative study, the average spectra of RT_{30} were processed in Table II. The average spectrum of each mosque group [i.e., Groups (A)–(F)] is presented in Fig. 7(a). The mid-frequency region of 500 and 1000 Hz, ($RT_{30\,m}$), usually provides a relative indication of the listening conditions in most rooms. The $RT_{30\,m}$ values ranged from 1.1 to 2.1 s with a mean value of 1.6 s and a STD of 0.4 s. The mid-frequency mean $C_{50\,m}$ is −1.4 dB, varying from −3.8 d to 0.8 dB with a STD of 1.7 dB. Figure 7(b) illustrates the C_{50} average spectra range as measured in all the sample mosques.

The measured RT_{30} of 20 mosques out of the total of 21 was found to be greater than the 1.0 s that is considered optimal upper limit for speech intelligibility. The average RT_{30} spectra depicted in Fig. 7(a) can be considered representative of all the sample mosques. It shows a long reverberation time at low frequencies (125–250 Hz) which is more than 45% of $RT_{30\,m}$ with high variations of 0.5 s at 250 Hz. This long sound decay at low frequencies can be detrimental for speech intelligibility. For good speech intelligibil-

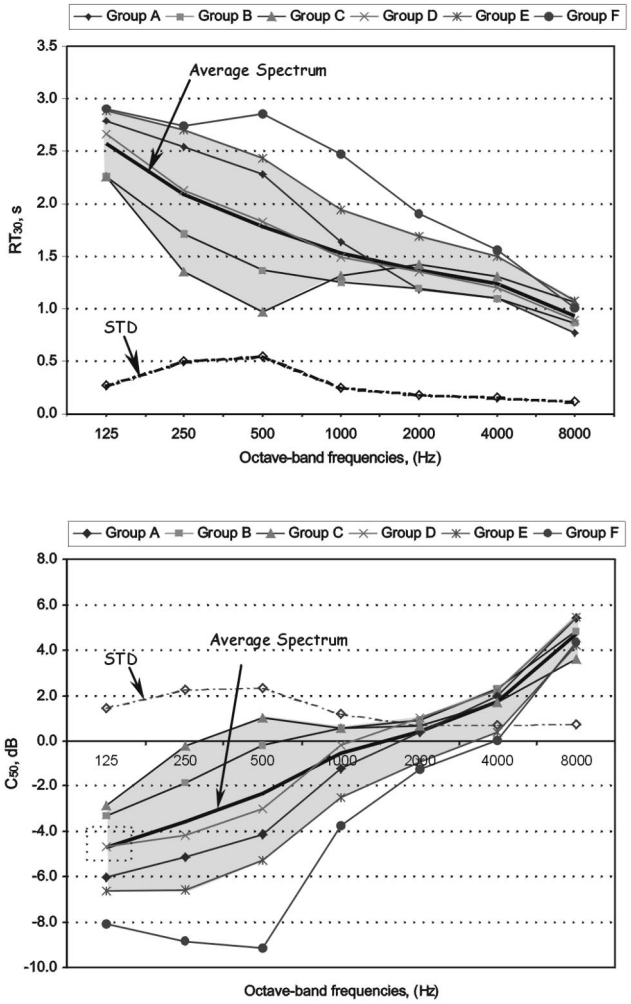


FIG. 7. Spatial average spectra of (a) RT_{30}, and (b) C_{50} in each mosque group compared to the value range, i.e., spatial minimum, maximum (shaded area), and average (——) of all groups except Group F.

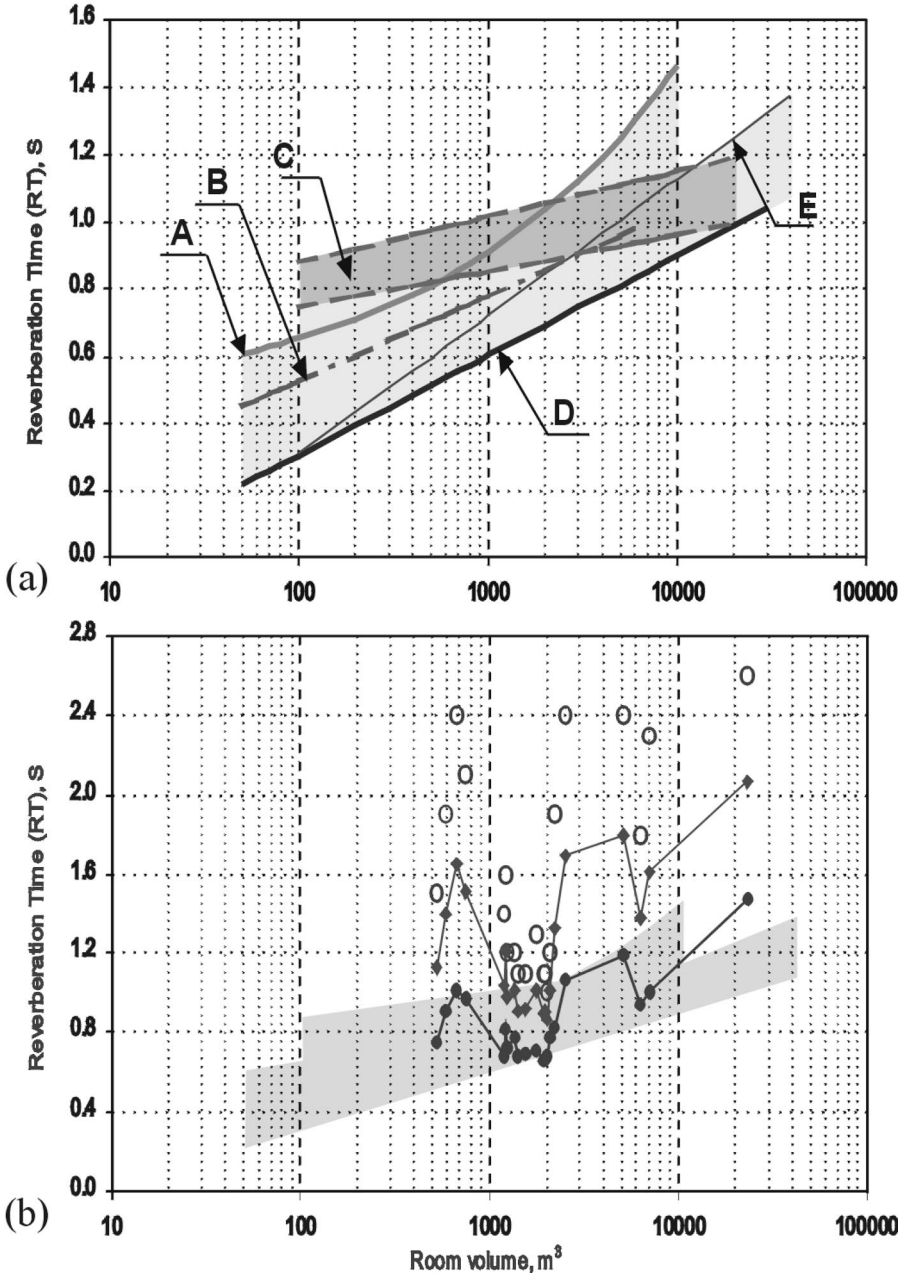


FIG. 8. (a) Recommended RT (optimal) for speech versus room volume (adapted from different sources), and (b) mid-frequencies (500–1000) RT values measured in all sample mosques ○ unoccupied, ♦ with 1/3 and ● 1/1 of the mosque's occupancy in comparison with suggested optimal values (shaded area). (A) In large rooms (Stephen & Bate equation) (Ref. 36, p. 34). (B) (At 500 Hz) (Ref. 36, Fig. 3.9, p. 36). (C) In conference rooms (Ref. 37, Fig. 2.29, p. 68). (D) (At an average of 500–1000 Hz) [Ref. 38, Fig. 26.25 (after E. B. Magrab), p. 1352]. (E) Maximum RT in large rooms [Ref. 38, Fig. 26.26 (after Bruel and Kjaer), p. 1353].

ity, RT values at low octave-band frequencies should remain flat down to 100 Hz.[35] An increase, at low frequencies of around 10%–20% of $RT_{30\,m}$ values would still be acceptable to yield a natural sound impression.

Various values and ranges have been proposed by many authors[36–39] for optimal RT values for speech. These were obtained and are presented in Fig. 8(a). The $RT_{30\,m}$ of the occupied mosques was then determined from measured $RT_{30\,m}$ in the unoccupied conditions. The $RT_{30\,m}$ values were calculated for 1/3, 2/3, and full mosque occupancy. Figure 8(b) depicts the measured $RT_{30\,m}$ (unoccupied) compared to $RT_{30\,m}$ in the three occupancy conditions compared with the several suggested optimal value ranges and trends. As can be seen, only a few mosques, particularly those in Group (C), were found with an $RT_{30\,m}$ close to the optimal RT range when unoccupied. The $RT_{30\,m}$ values of half of the sample mosques were greater than optimal, even with each mosque assumed to be 1/3 full, which is usually the case during performing *Daily* prayers. The $RT_{30\,m}$ of all the sample mosques came close to the boundaries or within the optimal RT range only with the mosques fully occupied, with the exception of mosque DM43 built with a huge volume as a landmark mosque intended for use by worshippers to perform *Friday* prayer only.

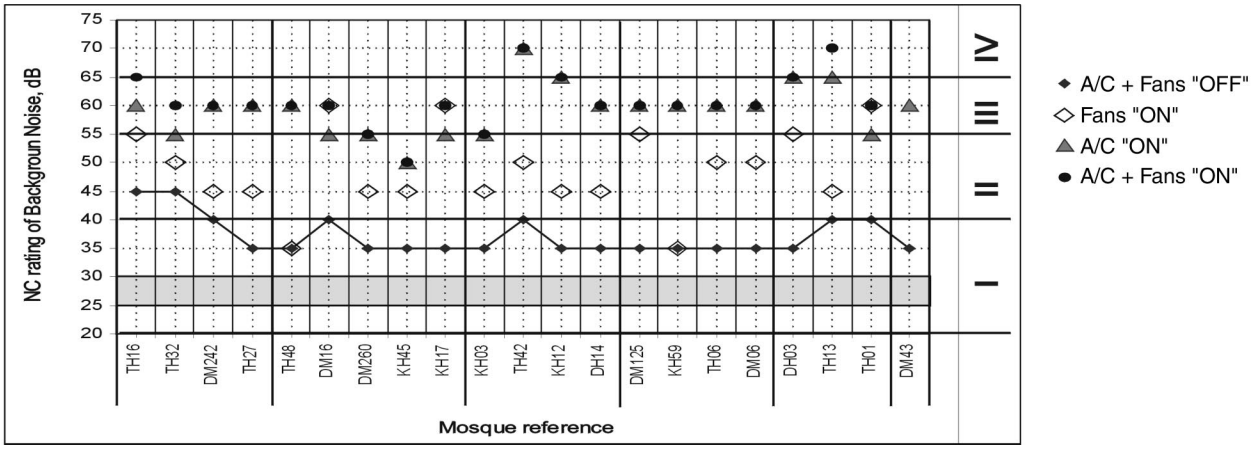


FIG. 9. A comparison of the BN and NC rating measured in all sample mosques at different operating conditions of fans and A/C systems against target ratings. I="very quiet" to "quiet," II="moderate noisy" to "noisy," III="very noisy," and VI="extremely noisy." (See Table III.)

B. Impact of air-conditioning and mechanical ventilation systems

To characterize the BN in mosques at different operating conditions, the spatial-average spectra of BN measured when fans and the A/C system were "OFF" are shown in Fig. 9. The spectra total A-weighted and linear values are also shown along with the NC ratings. It also includes the minimum, average, and maximum of mean spectra measured in all the sample mosques. Figure 10 depicts the BN average and value range compared to the average spectra measured in the three conditions of operating the active environmental control systems. The NC rating of measured BN ranged from NC-35 to NC-45. However, for very good speech listening conditions in spaces such as meeting rooms, conference halls, and courtrooms, NC is preferred to range from NC-25 to NC-30. Similarly, the maximum NC-30 criterion should not be exceeded because higher levels can be noticeable, and annoy many occupants (worshippers) or interfere with speech communication. The measurement of BN in mosques indicates a *moderately noisy* to *noisy* environment, but when the fans and A/C systems were operated the BN rating increased, resulting in a *very noisy* to *extremely noisy* environment. This can be expected to negatively affect speech intelligibility. Figure 10 compares measured BN at all selected mosques at different operation conditions with the NC-25, NC-30 range as a design goal.

C. Speech intelligibility and overall effectiveness of SRS

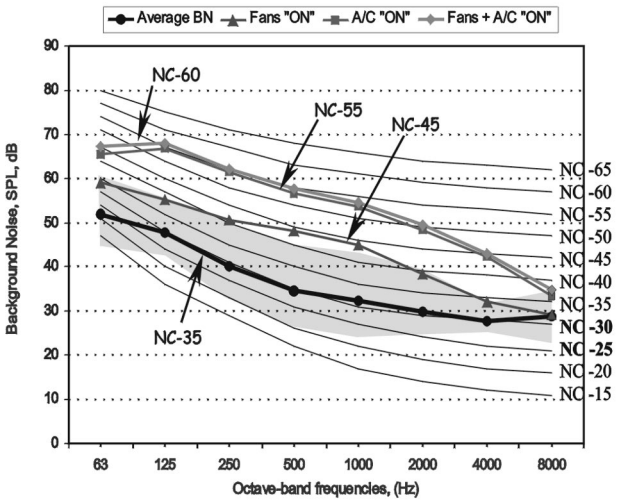
Following the measurements procedure motioned earlier, it was possible to characterize and rate speech intelligibility in the sample mosques. Figure 11(b) comprehensively compares speech intelligibility expressed by the STI. Average measured values are indicated by a vertical tick on each horizontal bar, which presents the value range of minima and maxima or all measurements conducted in each mosque. The number of measurement points is indicated along with the SI rating with and without the SRS being operated. As can be seen, the average SI ratings, for almost all mosques not operating the SRS, determined from STI values, lie in the range of *poor* to *fair* SI rating. The effect of using the SRS did improve SI in many, but not all, mosques. The effectiveness of many of the installed SRS can be clearly seen in the narrowing of the wide range of STI (without using the SRS) to a smaller value range (i.e., shortening the horizontal bar to narrower limits). This indicates that the SRS was effective in reducing the wide variation of SI to a more uniform value over the mosque floor area and improving SI. It must be indicated that the SRS were operated without the operation of fans and A/C systems, i.e., with BN rating ranging from NC35 to NC-45. Operating the fans and A/C system increased ambient BN to NC-60 and above. The effectiveness of the SRS in this case was not assessed and is questionable.

VI. CONCLUSIONS

Typical mosques of different sizes and shapes are usually designed considering the functional requirements with

TABLE III. Average BN measured in all sample mosques in comparison with NC curves at different operating conditions of fans and A/C systems.

	Overall SPL		NC			
Mosque reference	dB(A)	dB(L)	Average BN	FANS "ON"	A/C "ON"	A/C+FANS "ON"
TH16	45.4	63.3	45	55	60	65
TH32	48.7	62.7	45	50	55	60
DM242	42.9	56.8	40	45	60	60
TH27	39.1	54.7	35	45	60	60
TH48	39.4	57.0	35	35	60	60
DM16	45.2	65.9	40	60	55	60
DM260	40.5	60.4	35	45	55	55
KH45	39.4	51.1	35	45	50	50
KH17	37.7	59.5	35	60	55	60
KH03	36.8	54.6	35	45	55	55
TH42	42.9	61.1	40	50	70	70
KH12	36.6	57.1	35	45	65	65
DH14	34.9	53.8	35	45	60	60
DM125	36.6	58.4	35	55	60	60
KH59	38.3	56.0	35	35	60	60
TH06	40.3	60.4	35	50	60	60
DM06	36.6	58.4	35	50	60	60
DH03	37.5	54.7	35	55	65	65
TH13	40.1	56.8	40	45	65	70
TH01	44.2	63.0	40	60	55	60
DM43	39.3	57.1	35	NA	60	NA



	dB(A)	dB(L)	NC
Minimum	34.9	51.1	**30.0**
Average	40.1	58.2	**35.0**
Maximum	48.7	65.9	**45.0**
STD	3.5	3.6	
Fans "ON"	50.7	66.5	**45.0**
A/C "ON"	59.9	72.5	**55.0**
Fans + A/C "ON"	61.0	75.1	**60.0**

FIG. 10. The average BN measured in all sample mosques in comparison with NC curves at different operating conditions of fans and A/C systems. The shaded area represents the measured average value range.

regard to size and aesthetics, but with little attention paid to their acoustical quality. Acoustics in mosques are generally considered after mosques are built and involve mainly the process of providing sound system equipment. Audibility, not speech intelligibility is considered important and the acoustical considerations are not thought of either in the design phase or in the selection or installation of SRS in mosques. In this study, field measurements employing impulse response techniques were conducted to identify the acoustical characteristics and performance of selected mosques in Saudi Arabia. Representative sample mosques were objectively selected and investigated. A database of impulse responses captured in the sample mosques (unoccupied) utilizing impulse response measurements was established. The impulse responses were then analyzed to document and indicate the overall acoustical performance of the different mosque prototypes. Room-acoustic indicators including RT_{30}, C_{50}, BN, and STI were examined. Obtained results were compared to design goals. The following conclusions can be drawn.

$RT_{30\,m}$ relating to the mid-frequency region of 500 and 1000 Hz usually provides a relative indication of the listening conditions in most rooms. The $RT_{30\,m}$ of 20 mosques (unoccupied) out of a total of 21 was found to be greater than 1.0 s. The types of mosque investigated in this study can be characterized by long sound decay at low frequencies. This long sound decay can be detrimental, especially for speech intelligibility. For good speech intelligibility, RT values at low octave-band frequencies are preferred to remain flat down to 125 Hz. The $RT_{30\,m}$ of the occupied mosques was determined from measured $RT_{30\,m}$. The $RT_{30\,m}$ was calculated for 1/3, 2/3, and full mosque occupancy. Only a few mosques, particularly those of Group (C), possessed an $RT_{30\,m}$ close to the optimal RT range when unoccupied. The $RT_{30\,m}$ of half of the sample mosques was greater than optimal values even with the mosques 1/3 full. The $RT_{30\,m}$ of all sample mosques (except mosque DM43) approached the boundaries came within the optimal RT value range only with the mosques fully occupied. The recommended volume (m^3/occupant) for rooms intended for speech is 3.0 m^3 up to a maximum of 5.0 m^3. Volumes per worshipper in all sample mosques ranged from 4.2 to 5.0 m^3 expect the mosque DM43, which had an 8.3 m^3 per worshipper, exceeding the recommended value range.

In all mosques, the NC ratings obtained from measured BN range from NC-35 to NC-45. In the literature there is no explicit preferred NC range specified for mosques. However, for very good speech listing conditions in spaces such as meeting rooms, conference halls, and courtrooms NC is preferred to range from NC-25 to NC-30. The cause of high NC ratings obtained from measurements can be attributed to intruding exterior noise due to low transmission loss of exterior wall systems, particularly windows and glazing types commonly installed. An additional factor includes interior noise such as buzzing and humming from defective lighting unit accessories. It is recommended that the NC-30 criterion should not be exceeded because higher levels can be noticeable and annoy most worshippers as well as interfere with speech communication. The measurement results of BN in mosques indicated a moderately noisy to noisy mosque environment. The measurements also showed that when components of active environmental control systems such as fans and A/C units were operated the NC rating increased, resulting in a very noisy to extremely noisy environment. This can be expected to adversely affect speech intelligibility. Care should therefore be taken to select and install or mount A/C system units with low noise output and lighting units with high quality ballast and accessories that require minimal maintenance. The mosque exterior wall should be designed to minimize intruding exterior noise to satisfy the NC-25 criterion.

The average SI ratings resulting from measurements in the sample mosques without operating the SRS determined from STI values, were found to be in the range of poor to fair SI. The use of SRS did improve SI in many but not in all cases. Many of the installed SRS in mosques contributed to a reduction in the variability of SI. This indicates that SRS were generally effective in reducing the wide variation of SI to a more uniform one over the mosque floor area, particularly at more remote locations away from the *Imam's* position and slightly improving SI. It must be indicated that the SRS were investigated with fans and A/C units set to the ''OFF'' condition (as is the case in the autumn, winter, and spring conditions), resulting in an average ambient BN with NC-40 rating. Operating either the fans or the A/C system only, or both, proved to increase ambient BN to NC-60 and above. The effectiveness of the SRS in this case was not assessed and is questionable. The developed database can be utilized as a valuable source of objective acoustical indicators and comparative information for further investigations. Research work is needed to characterize the frequency-

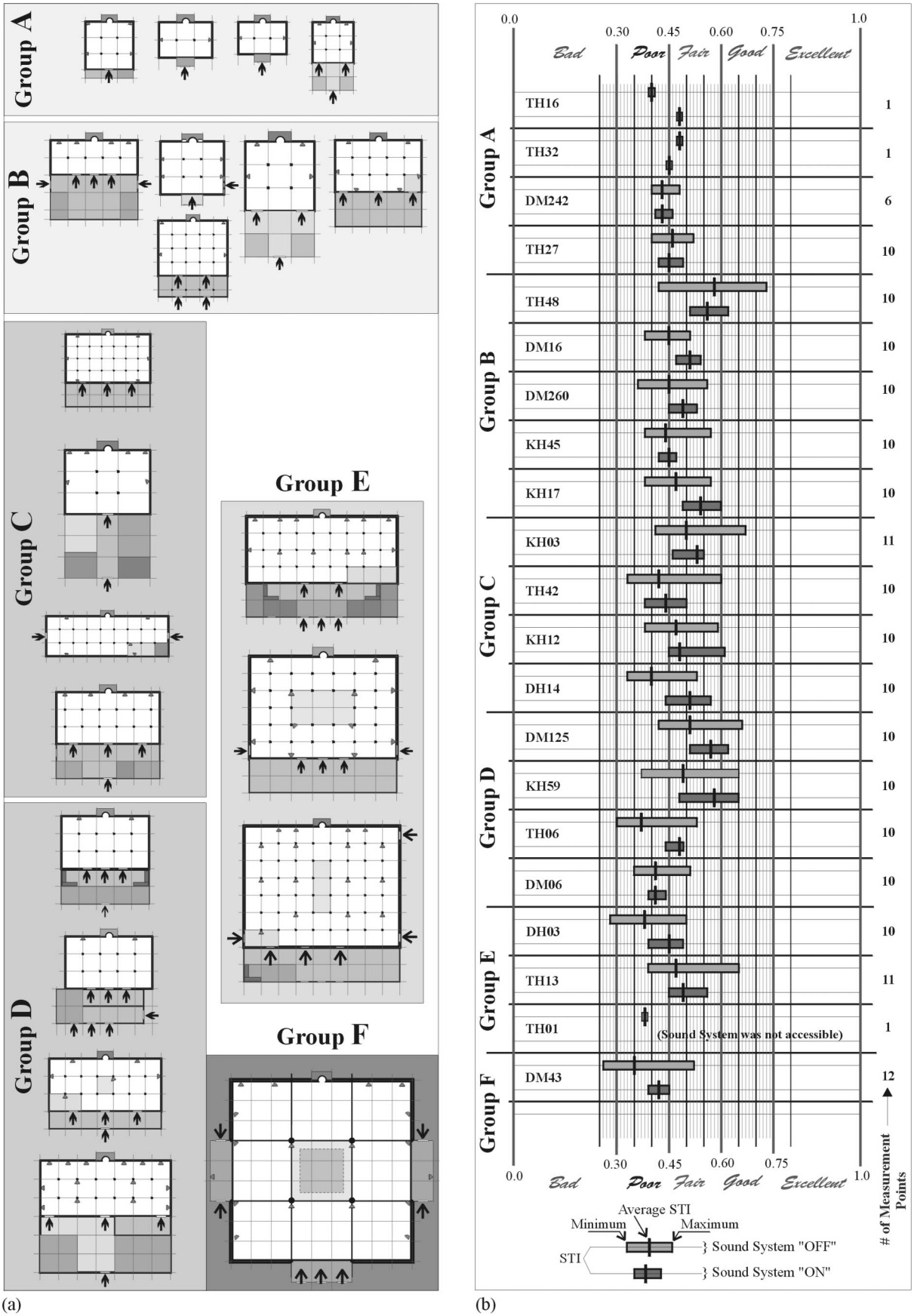


FIG. 11. (a) Grouping of the selected sample mosques based on volume and congregational capacity of prayer-performing hall (a relative scale is used). Note: The distribution of loudspeakers installed in each mosque is shown, and (b) a comparison of STI measured in all sample mosques with and without operating the mosque's SRS. The target value range is highlighted. (Note: mosques are arranged in ascending order with respect to volume.)

dependent absorption coefficients of worshippers in mosques in the different worship modes. In the mosque design phase, this would help architects to reasonably estimate, with some degree of accuracy, the reverberation time with varying occupancy compared to recommended value ranges. In addition, the use field measurements, such as those reported here, in combination of utilizing computer model studies can be an effective approach for developing better information for designing better mosques from the acoustical viewpoint.

ACKNOWLEDGMENTS

The author would like to express his gratitude and thanks to *King Abdulaziz City for Science and Technology (KACST)* for funding this research project and facilitating its implementation. Thanks are also due to *King Fahd University of Petroleum and Minerals (KFUPM)* for its continuous support and encouragement. Gratitude is also extended to the local authorities of the *Ministry of Islamic Affairs, Endowments, Da'wah and Guidance* for their cooperation and for the provision of information. The efforts of the University Editorial Board are appreciated.

[1] R. N. Hammad, ''RASTI measurements in mosques in Amman, Jordan,'' Appl. Acoust. **30**, 335–345 (1990).
[2] M. K. Abdelazeez, R. N. Hammad, and A. A. Mustafa, ''Acoustics of King Abdullah mosque,'' J. Acoust. Soc. Am. **90**, 1441–1445 (1991).
[3] A. El Bashir and A. Al-Gunaimi, ''Evaluation of acoustical performance of mosques: a case study on a typical mosque in the Gulf region,'' *Proceedings of the 4th Saudi Engineers Conference*, Kingdom of Saudi Arabia, Jeddah, 1995, Vol. 1.
[4] A. Abdou, ''Predicting and assessing the acoustical performance of mosques employing computer simulation: *a case study*,'' *Proceedings of the Symposium on Mosque Architecture*, 1999, Vol. 6, pp. 73–87.
[5] D. L. Klepper, ''The distribution column sound system at Holy Cross Cathedral, Boston, the reconciliation of speech and music,'' J. Acoust. Soc. Am. **99**, 417–425 (1996).
[6] J. Anderson and T. Jacobsen, ''RASTI measurements in St. Paul's Cathedral,'' London, Brüel & Kjaer Note No., 1985.
[7] R. W. Guy and A. Abdou, ''Acoustic quality via sound intensity measurements in a church,'' *Proceedings of Inter. Noise 94*, Yokohama, Japan, 1994, pp. 1851–1854.
[8] A. P. O. Carvalho, ''Relationships between objective acoustical measures and architectural features in churches,'' *Proceedings of the W. C. Sabine Symposium*, the 127th Meeting of the Acoustical Society of America, 1994, pp. 311–314.
[9] P. O. Carvalho, A. E. J. Morgado, and L. Henrique, ''Relationship between speech intelligibility and objective acoustical parameters for architectural features in catholic churches.'', J. Acoust. Soc. Am. **101**, Pt 2, 3051(A) (1997).
[10] M. Galindo, T. Zamarreno, and S. Giron, ''Clarity and definition in Mudejar-Gothic churches,'' J. Building Acoust. **6**(1), 1–16 (1999).
[11] M. Barron, ''Impulse testing techniques for auditoria,'' Appl. Acoust. **17**, 165–181 (1984).
[12] H. Tachibana, H. Yano, and Y. Hidaka, ''Measurement of impulse response and its applications in room acoustics,'' J. Acoust. Soc. Am. **89**, 1856 (1991).
[13] G. R. D. King, *The Historical Mosques of Saudi Arabia* (Longman Group Ltd., United Kingdom, 1986).
[14] I. Sergeld and J. Steele, in *Architecture of The Contemporary Mosque*, edited by Ismail Serageldin and James Steele (Academy Group Ltd., London, 1996).
[15] H. M. Ibrahim, *Planning Standards for Mosques* (in Arabic), Ministry of Municipal & Rural Affairs, Kingdom of Saudi Arabia, 1979.
[16] Documents of the Ministry of Islamic Affairs, Endowments, Da'wah and Guidance (*MIAEGD*), Kingdom of Saudi Arabia.
[17] A. Abdou and R. W. Guy ''Spatial information of sound fields for room-acoustic evaluation and diagnosis,'' J. Acoust. Soc. Am. **100**, 3215–3322 (1996).
[18] J. P. Lochner and J. F. Burger, ''The influence of reflections on auditorium acoustics,'' J. Sound Vib. **1**, 426–454 (1964).
[19] J. S. Bradley, ''Predictors of speech intelligibility in rooms,'' J. Acoust. Soc. Am. **80**, 837–845 (1986).
[20] J. S. Bradley, ''Speech intelligibility studies in classrooms,'' J. Acoust. Soc. Am. **80**, 846–854 (1986).
[21] T. Houtgast and H. J. M. Steeneken, ''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 (1985).
[22] H. J. M. Steeneken and T. Houtgast, ''RASTI: a tool for evaluating auditoria,'' Technical Review, Brüel & Kjaer, No. 3, 1985, pp. 13–30.
[23] ''The objective rating of speech intelligibility in auditoria by the 'RASTI' method,'' IEC 268, part 16, Sound System Equipment, 1988.
[24] V. M. A. Peutz, ''Articulation loss of consonants as a criterion for speech transmission in a room,'' J. Audio Eng. Soc. **9**, 915–919 (1971).
[25] J. S. Bradley, ''Relationships among measures of speech intelligibility in rooms,'' J. Audio Eng. Soc. **46**, 396–405 (1998).
[26] P. W. Barnett, ''A review of speech intelligibility indicators—their relationship and applications,'' J. Acoust. Soc. Am. **101**, No. 5, 3050(A) (1997).
[27] J. van der Werff, ''Speech intelligibility in rooms,'' *Proceedings of Inter-Noise 97*, 15–17 June 1997, pp. 349–354.
[28] M. Hodgson, R. Rempel, and S. Kennedy, ''Measurement and prediction of typical speech and background noise levels in university classrooms during lectures,'' J. Acoust. Soc. Am. **105**, 226–233 (1999).
[29] M. Hodgson, ''Experimental investigation of the acoustical characteristics of university classrooms,'' J. Acoust. Soc. Am. **106**, 1810–1819 (2000).
[30] S. R. Bistafa and J. S. Bradley, ''Reverberation time and maximum background-noise level for classrooms from a comprehensive study of speech intelligibility metrics,'' J. Acoust. Soc. Am. **107**, 861–875 (2000).
[31] D. D. Rife, ''Maximum length sequence system analyzer,'' Reference Manual, DRA Laboratories, 1987–1996, p. 86.
[32] ISO 3382, ''Acoustics measurement of reverberation time of rooms with reference to other acoustical parameters,'' 1997.
[33] H. J. M. Steeneken and T. Houtgast, ''A physical method for measuring speech-transmission quality,'' J. Acoust. Soc. Am. **61**, 318–326 (1980).
[34] N. R. French and J. C. Steinberg, ''Factors governing the intelligibility of speech sounds,'' J. Acoust. Soc. Am. **19**, 90–119 (1947).
[35] H. Kuttruff, *Room Acoustics*, 3rd ed. (Elsevier, London and New York, 1991), p. 164.
[36] B. J. Smith, R. J. Peters, and S. Owen, *Acoustics and Noise Control* (Longman, London/New York, 1988), pp. 34–36.
[37] D. Templeton, P. Sacre, P. Mapp, and D. Saunders, *Acoustics in the Built Environment*, 2nd ed. (Architectural Press, City, 1997), pp. 58–68.
[38] B. Stein and J. S. Reynolds, *Mechanical and Electrical Equipment for Buildings*, 8th ed. (Wiley, New York, 1992), p. 1325.
[39] M. D. Egan, *Architectural Acoustics* (McGraw-Hill, United Kingdom, 1998), p. 370.

Phase change measurement, and speed of sound and attenuation determination, from underwater acoustic panel tests

Jean C. Piquette[a] and Anthony E. Paolero
Naval Undersea Warfare Center, Division Newport, 1176 Howell Street, Newport, Rhode Island 02841

(Received 14 March 2002; revised 28 October 2002; accepted 4 November 2002)

A technique for measuring the change in phase produced by the insertion of a panel between a projector and receiver is described. Presented also is a procedure for determining the phase speed and attenuation of the panel material. Although the methods were developed over the frequency decade 10–100 kHz, they are not limited to that band. It was observed that a ''settling time'' of approximately 20 min is required to obtain reproducible phase measurements if the experiment is disturbed even slightly. For example, rotating the panel 10 degrees, then immediately returning to the original position, causes the observed phases to differ by up to 10 deg from those obtained prior to the disturbance. These differences are distributed randomly across frequency. Temperature stabilization within the medium as well as the material is also required before measurements can take place. After the stated 20 min settling time, however, the phases return to the values obtained prior to rotation, or after temperature stabilization, to within $\pm\frac{1}{2}$ deg. The sound speed and attenuation determination technique employs least-squares fitting of a causal model to the measurements. Four (or fewer) adjusted parameters accommodate the measurements over the stated frequency decade, even for samples that exhibit significant dispersion. The sound speed is typically determined to an accuracy of ± 30 m/s, as judged from a propagation-of-error calculation. This model assumes single-layered panels.
[DOI: 10.1121/1.1531984]

PACS numbers: 43.58.Dj, 43.58.Vb, 43.60.Qv [SLE]

I. INTRODUCTION

An ideal acoustic window would have acoustical properties that are identical to those of the surrounding medium. If a sample panel of such an ideal window material is inserted between a projector and receiver, no change in the amplitude or phase of the signal detected by the receiver would occur with respect to that observed prior to the insertion. While measurement of the change in wave *amplitude* is easily done using standard methods,[1] determination of the change in wave *phase* is usually avoided, owing to the difficulty of performing the measurement accurately. As test frequency is increased, the measurement becomes increasingly difficult. The increase in difficulty is caused by the need to ensure that little or no displacement occurs in the positions of any of the acoustic elements, and that no temperature change occurs in the surrounding acoustic medium. For example, at a test frequency of 100 kHz a change in the distance between the projector and the receiver of just 1 mm introduces an error in the phase-change measurement of approximately 24 deg for a water medium at room temperature, thus invalidating the measurement. If the test procedure necessitates moving the acoustic elements in order to permit insertion and removal of the panel, as is usually the case in typical underwater acoustic panel tests, relative displacements of the acoustical elements exceeding 1 mm are virtually assured.

The measurement of the phase change caused by the insertion of a panel between a projector and receiver can be used to help determine the sound speed and attenuation of the sample material, assuming a single-layer homogeneous panel whose thickness and density are accurately known. For example, an acoustic window of 1-cm thickness whose sound speed differs from that of the surrounding water medium by only 10 m/s will produce a phase change of about 1.6 deg at 100 kHz when inserted between source and receiver. A simple causal model of the acoustical behavior of the material can be fitted, in the least-squares sense, to the phase-change measurements acquired across a broad frequency band. In this way it is possible to deduce both the sound speed and the attenuation of the acoustic material under test through the adjustment of only a very small number of model parameters. In the model considered, four parameters (or fewer) have been found to be sufficient to accommodate measurements from acoustic windows over the frequency decade of 10–100 kHz.

In Sec. II a description of the experimental setup and procedure is given. Section III presents experimental results, and provides a comparison of measured phase changes from a steel plate to the phase changes expected theoretically. Considered are incidence angles from normal to 40 deg off the normal. The method for phase speed and attenuation determination is described in Sec. IV. Applications of that method to data are presented there as well. A summary and the conclusion are given in Sec. V.

II. EXPERIMENTAL METHOD

The measurements are conducted in a large test vessel (see Fig. 1), the Acoustic Pressure Tank Facility (APTF) of the Naval Undersea Warfare Center, Division Newport, lo-

[a]Electronic mail: piquettejc@npt.nuwc.navy.mil

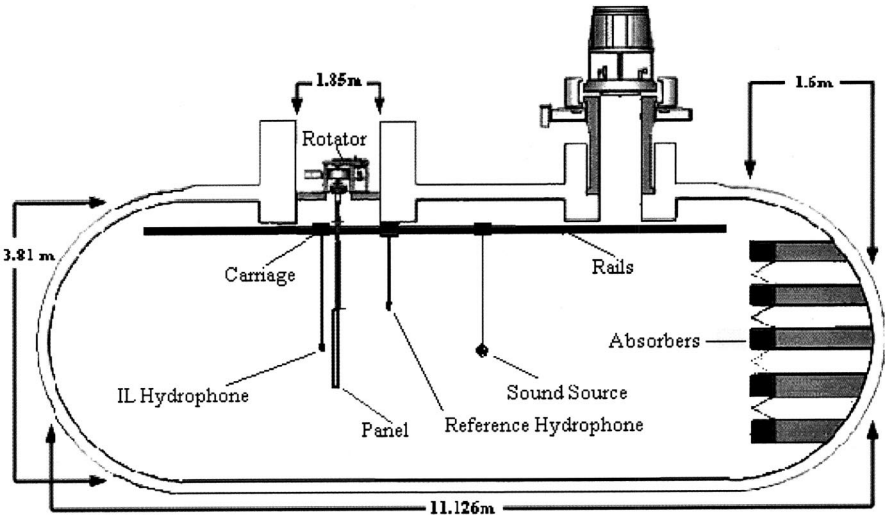


FIG. 1. The test vessel. The measurement configuration depicted permits the panel to be inserted into, and removed from, the vessel without requiring movement of any of the acoustical elements. The configuration also allows for the panel to be rotated up to 80 deg off of the normal-incidence condition without moving of any acoustical elements. Here, ''IL hydrophone'' denotes the insertion loss hydrophone, as described in the text.

cated in Newport, Rhode Island. This tank is capable of achieving hydrostatic test pressures up to 18.6 MPa (2700 psi) and test temperatures from 2 to 40 °C, although the present work was limited to measurements acquired at atmospheric pressure. An important feature of this facility for the present method is the three-carriage underwater positioning system, as well as the underwater rotator located under the larger of the two access ports. This rotator can accommodate panels up to 2700 kg in mass. Highly accurate stepper motors control the carriage positioning system and the rotator angle.

The measurement configuration includes a USRD F42A (spherical) projector and two USRD H52 hydrophones.[2] The F42A source was selected because it is capable of covering the 10–100-kHz frequency band of interest, and because its spherical design permits uniform insonification of the sample as well as uniform insonification of the two hydrophones used in the procedure. The source is positioned in the $X-Z$ plane configuration, relative to the hydrophones, in order to uniformly insonify the hydrophones within the source's equatorial plane. One of the two hydrophones, termed the ''insertion loss (IL) hydrophone,'' is situated ''behind'' the sample panel and, together with the sound source, defines the acoustic axis of the measurement. This is the usual location of a hydrophone in an insertion loss measurement.[1] The second hydrophone, termed the ''reference hydrophone,'' is located in ''front'' of the sample panel, and is located well off of the acoustic axis in order to avoid acoustic ''shadowing.'' Its purpose is to monitor the incident sound field, to ensure its constancy from one measurement to the next. (The sound field measured by the reference hydrophone is gated in time to avoid reflections from the panel, so that this hydrophone always acquires data under free-field conditions whether or not a sample panel is present.) Should the sound field observed at the reference hydrophone change between measurements acquired with the panel present and those acquired with the sample absent, it is sometimes possible to use the measurements acquired by the reference hydrophone to correct those acquired by the insertion loss hydrophone. For example, since the projector, insertion loss hydrophone, and reference hydrophone are never disturbed at any point during the test, a correctable phase error may be attributed to a change in water temperature that may occur between measurements. Because the temperature of the facility workspace is always maintained at a comfortable working level, the water in the tank unavoidably changes temperature during tests in which the water has been heated or chilled. That is, the water slowly cools when it has been heated above the temperature of the facility atmosphere, and slowly warms when it has been chilled below the temperature of the facility atmosphere. But, this effect is clearly revealed by subtracting phase measurements acquired at the reference hydrophone from run to run, and this reference calculation can be used to correct the phase measurements acquired at the insertion loss hydrophone. (As can be seen from Fig. 1, the reference and insertion loss hydrophones are rigged relatively close to each other, so that each hydrophone can reasonably be expected to experience essentially the same change in water temperature. Although the two hydrophones are rigged at slightly different depths, this was not found to cause a problem in the experiments. This is likely due to the fact that insertion and removal of the sample panel sufficiently mixes the water.)

III. EXPERIMENTAL RESULTS

A. Measurement reproducibility and "settling time"

Measurement reproducibility tests were done over intervals of days in order to check the stability and reliability of the method. The tests were conducted over the frequency interval of 10–100 kHz, at test temperatures of 10, 20, and 30 °C. A simple steel plate was used as a test sample during this part of the investigation to help determine, through comparison of theory and experiment, whether phase change measurements could be accurately performed. It was found that phase could be accurately, and reproducibly, measured over the specified frequency band to within an accuracy of $\pm\frac{1}{2}$ deg. Water temperature was monitored with a digital thermocouple, and phase changes at the reference hydrophone consistent with the recorded water temperature changes were observed. (The water temperature changes that occur due to contact of the water with the facility atmosphere are typically no more than a few tenths of a degree centi-

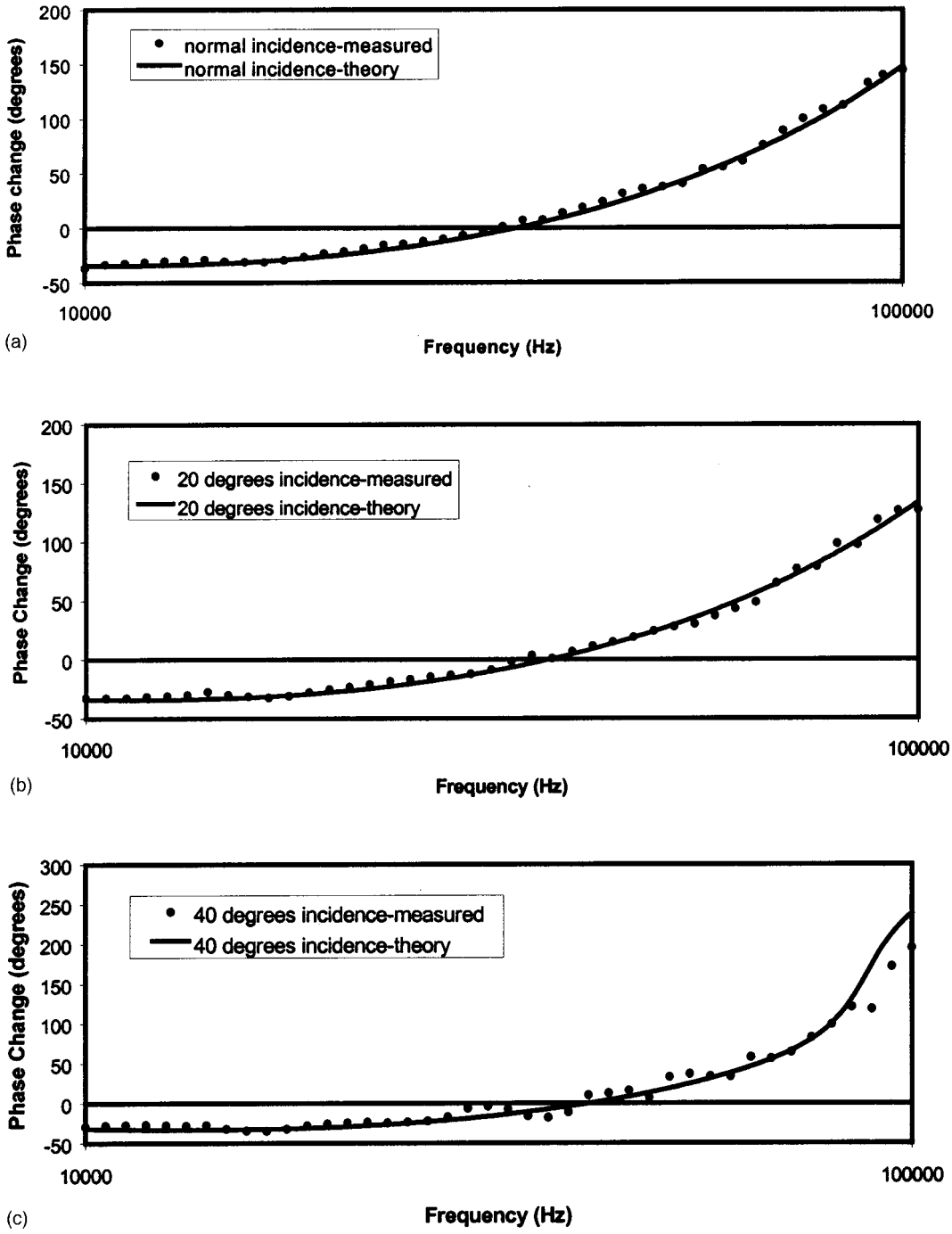


FIG. 2. Comparison of measured phase changes (dots) with theoretical phase changes (solid lines) for a steel plate at various angles of incidence. (a) Normal incidence. (b) 20 degrees incidence. (c) 40 degrees incidence.

grade over the standard workday during which measurements are ordinarily acquired.) These observations confirmed that it is possible to correct measurements obtained over a significant time interval (several hours) if a second-order difference method is used. That is, if a slope appeared in the graph of the differences between two phase versus frequency measurements as acquired at the reference hydrophone in frequency sweeps conducted several hours apart, a similar slope was also observed in the plot of the differences obtained at the insertion loss hydrophone. Hence, the differences between the two reference hydrophone curves could be used to correct the corresponding differences seen at the insertion loss hydrophone, after accounting for any difference in distances of the reference phone and IL phone from the source. Such a capability is important, because when a sample is present between the projector and the receiver all changes in phase seen at the insertion loss hydrophone would otherwise be attributed to the influence of the material of the panel.

A very significant effect was observed with respect to motions induced in any of the acoustical elements or the sample. For example, once a desired measurement configu-

ration was established, free-field data (panel absent) were acquired at both the reference and insertion loss hydrophones. The carriage holding the insertion loss hydrophone was then displaced along the rail system (see Fig. 1), using the highly accurate stepper-motor control system, so that the distance between the projector and the insertion loss hydrophone increased by 15 cm. The carriage was then immediately returned to its original position, and free-field data were newly acquired from both the reference hydrophone and the insertion loss hydrophone. While the phases obtained at the reference phone were found to be unchanged with respect to those observed prior to the motion of the insertion loss hydrophone, the phases obtained at the insertion loss hydrophone exhibited large differences with respect to those acquired by it previously. These differences were randomly distributed in frequency. The differences were found to be as great as 10 deg at various frequencies. A frequency sweep to acquire phase data from both hydrophones was then conducted every 5 min. It was found that the differences between the phases observed at the insertion loss hydrophone from those acquired prior to moving that hydrophone continuously declined with time. After 20 min ''settling time,'' the phases observed at the insertion loss hydrophone differed from those observed prior to moving it by approximately $\pm\frac{1}{2}$ deg, with the remaining differences randomly distributed across the frequency band. (During this 20-min period, the water temperature changed so negligibly that no change was observed on the monitoring thermocouple, which can detect changes of about 0.1 °C. Typically, several hours must elapse in order for a measurable change in water temperature to occur.)

This 20-min settling time was found universally to be required after any disturbance of the sample or acoustical elements in order to obtain reproducible phase measurements. In another test that further illustrates the phenomenon, the phase was captured at the insertion loss hydrophone with the steel-plate sample in place. Then, the steel plate was lifted up close to, but not through, the water surface. The steel plate was then returned to its original location. The $\pm$10-deg randomly distributed phase excursions with respect to those obtained prior to moving the steel plate were again observed. However, after the 20-min settling time had elapsed, the measured phases once again were found to agree with those obtained prior to moving the plate, to within the $\pm\frac{1}{2}$-deg accuracy previously mentioned. Moreover, if the panel was rotated away from and back to a randomly selected reference angle, a 20-min settling time was again found to be needed before reproducible phase measurements could be obtained. A similar effect was observed for movements of the sound source. It is not known why a settling time of this duration is required, or what physical effect is causing these randomly distributed phase errors.

It was also found that, despite the high accuracy of the stepper-motor-driven carriage positioning system, it is not possible to move a carriage to which a hydrophone is rigged if accurate phase measurements are desired. In one set of measurement reproducibility tests, in which only a source and a single receiver were used, phase measurements were captured from a hydrophone suspended from a carriage. The carriage was then translated 15 cm away from its initial position, then translated back to its original position, and phase measurements were acquired. This reproducibility test was conducted a dozen times. It was found in most cases that the phase measurements would accurately repeat to within $\pm\frac{1}{2}$ deg of phase error across the frequency band of 10–100 kHz, after the required 20-min settling time. Then, after several such successful reproducibility tests, and for no apparent cause, the difference between the just-acquired phase data and the original (undisturbed) phase data, plotted vs frequency, would be found to abruptly exhibit a significant slope. (That is, the phase errors were not randomly distributed with frequency in this case, but rather increased linearly with frequency.) The slope was such that the newly acquired phase at 100 kHz differed by as much as $\pm$8 deg from the original value obtained at this frequency, and this difference did not decline with time. This result suggests a repositioning error in the hydrophone of approximately 0.33 mm. It was concluded that a carriage translation of any kind would be unacceptable in an experiment in which reproducible phase measurements are required. It is thus necessary to capture both the ''with panel'' and ''without panel'' data without moving, at any time, any of three carriages involved in the measurements.

B. Method validation by comparing theory and experiment for a steel plate

As a check of the validity of the experimental procedure, phase-change measurements were acquired from a flat steel plate of $76.2\times76.2\times0.95$ cm ($30\times30\times3/8$ in.) geometric dimensions. The phase change as defined here was experimentally deduced by first measuring the phase at both the insertion loss and reference hydrophones with no panel present. The phases were then measured with the panel in place. All measurements are then conducted a second time. That is, the panel is lifted to (but not through) the water surface, then replaced, and the phases are remeasured, so that measurement reproducibility information is available. (While the panel is situated near the water surface an additional sweep of free-field, or ''without panel,'' measurements are acquired, also for the purposes of checking measurement reproducibility of the incident sound field.) The phases measured by the reference hydrophone with the panel present are then subtracted, frequency-by-frequency, from those obtained by the reference hydrophone with the panel absent. These differences computed from the measurements obtained from the reference hydrophone are then subtracted, frequency-by-frequency, from the corresponding differences obtained from the insertion loss hydrophone. This second-order difference is taken to be the experimentally measured phase change caused by the insertion of the panel between the projector and receiver (where ''receiver'' refers to the insertion loss hydrophone). This procedure assumes that subtracting the differences obtained from the reference hydrophone from those obtained at the insertion loss hydrophone eliminates any changes in phase that are not associated with the influence of the material upon the sound field, but which are purely attributable to measurement error. Agreement be-

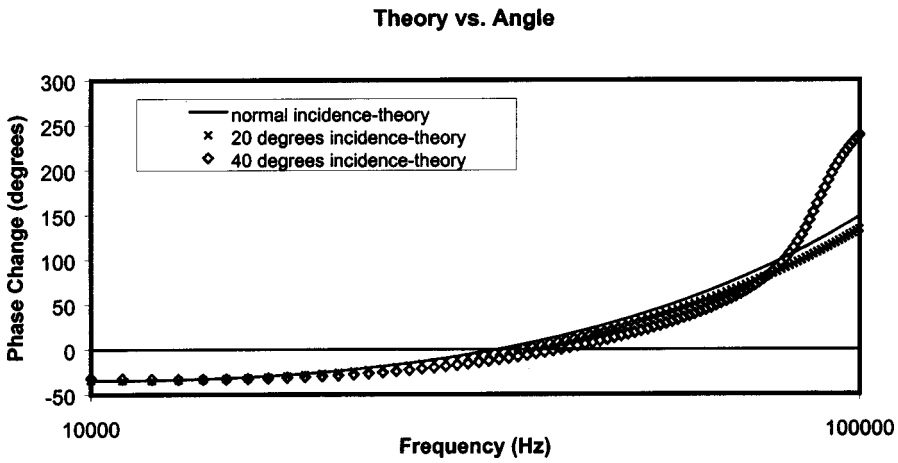


FIG. 3. Comparison of theoretical curves for various angles of incidence for a steel plate.

tween the two sets of phase-change measurements that are always conducted is taken to be a good indication that the measurement has been accurately conducted.

The results of these measurements for the case of the steel plate are summarized in Figs. 2(a)–(c). Here, the theoretical curves (solid lines) have been computed using a standard layer model,[3] and are based upon tabulated values for the acoustical properties of steel. (Since highly oblique angles of incidence were considered in the experiment, the layer model used accounts for the production of shear waves in the sample. However, the theoretical model ignores edge effects.) As can be seen, agreement between theory and experiment (dots) is quite good throughout, with accuracy declining at the highest frequencies in the 40-deg incidence case. It should be noted, though, that for a sample of the given dimensions such a high angle of incidence is nearing the limit of what can be measured in the absence of edge effects, even considering only the time of flight of the edge-diffracted wave through the water. It is possible that complex surface waves, generated at the sample edges and reradiated into the fluid from the flat surface of the plate, are affecting the higher-frequency results. If that is the cause of the differences between theory and experiment seen for the largest incidence angle, a panel that, unlike steel, has even a modest intrinsic attenuation would not be as significantly affected as is suggested in this measurement. The samples of actual measurement interest typically have at least some intrinsic attenuation, so the steel plate represents a particularly difficult case to test the method.

In order that the differences between the phase changes for each angle of incidence are more readily seen, the theoretical curves for each of the angles considered in Fig. 2 are compared in Fig. 3.

IV. DETERMINATION OF SOUND SPEED AND ATTENUATION

A. Theory

Since it was desired to use the phase-change measurement capability as the basis of a sound speed and attenuation

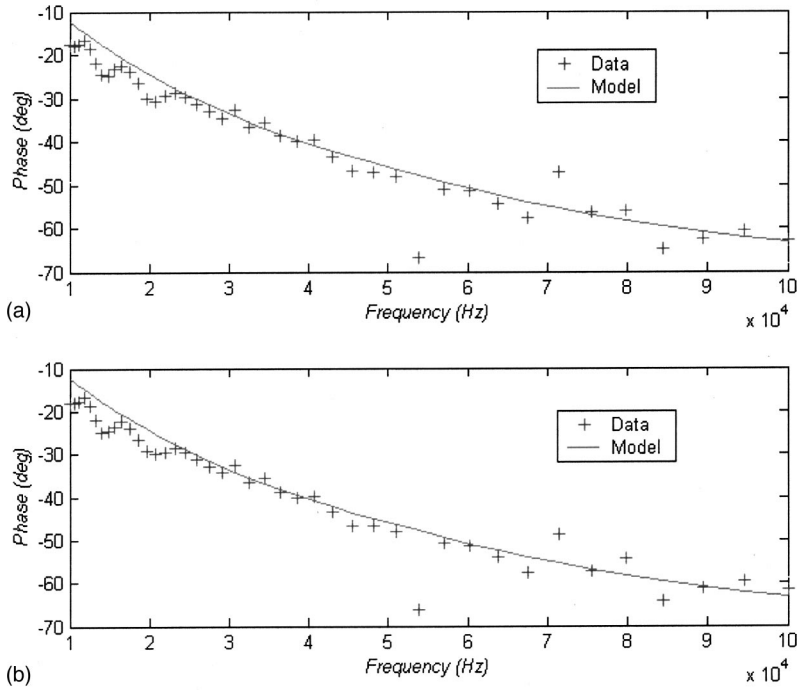


FIG. 4. Phase change for a dispersive sample. Solid lines are the results of the fit, and plus (+) symbols denote data. (a) First measurement. (b) Repeat measurement.

determination method, a causal theoretical model relating these material properties was derived. The model is known to be causal because it is compatible with the relevant Kramers–Kronig relationship. (See, for example, Ref. 4). The model parameters (of which there are four) are determined by least-squares fitting of the material model, as incorporated into a layer model, to the measured phase changes. In addition, the theoretical model is least-squares fitted to the usual insertion loss[1] that can also readily be deduced from the measurements acquired by the IL hydrophone.

The model considered here is similar to that used in the ONION method.[5] However, unlike that model, the present model is fully causal. Also unlike the ONION method, the present method uses phase-change data as well as the insertion loss data that are used in the ONION method. Moreover, the present method does not use, or require, the echo reduction.[1]

The model used here represents the attenuation in terms of a Maclaurin series truncated at the quadratic term. In order to avoid the use of specialized methods[6] in the calculation of the Hilbert transforms required for ensuring consistency with the Kramers–Kronig relationship, the Maclaurin series is subjected to a filtering function that forces the attenuation to rapidly approach zero beyond an arbitrarily selected cutoff frequency. The cutoff frequency is chosen well outside the frequency band of measurement interest. Thus, the model attenuation function assumes the form

$$\alpha(\omega)=\frac{a_0+a_1\left(\frac{\omega}{\omega_0}\right)+a_2\left(\frac{\omega}{\omega_0}\right)^2}{\left[1+b\left(\frac{\omega}{\omega_0}\right)^4\right]}. \tag{1}$$

Here, a_0, a_1, and a_2 are the (unknown) Maclaurin series expansion coefficients, which have the dimension of reciprocal distance. The quantity ω is the angular drive frequency and ω_0 is the (fixed) angular cutoff frequency (both expressed in radians/s) of the filter. Finally, the quantity b is a dimensionless parameter that provides some flexibility in setting the filter characteristics. (The cutoff frequency ω_0 can be used to entirely set the filter characteristics, so that the dimensionless filter parameter b is merely a convenience. Also, the use of a fourth-order filter here is arbitrary, although this filter was found in the present work to provide good computational properties.)

Using the appropriate Kramers–Kronig relationship,[4] the Hilbert transform of the attenuation function $\alpha(\omega)$ can be related to the speed of sound $c_{\text{phase}}(\omega)$ by the equation

$$\frac{\omega}{c_{\text{phase}}(\omega)}=\frac{\omega}{c_\infty}+[\alpha(\omega)]. \tag{2}$$

In Eq. (2), the bracket notation [ ] is used to denote the Hilbert transform, and the symbol c_∞ denotes the phase speed at infinite angular frequency. (The quantity c_∞, together with a_0, a_1, and a_2, completes the four-parameter set that is determined by least-squares fitting to the phase change and insertion loss data.) Carrying out the indicated Hilbert transform in Eq. (2) and solving the resulting equation for $c_{\text{phase}}(\omega)$ gives

$$c_{\text{phase}}(\omega)=\frac{2b^{1/4}c_\infty\pi(b\omega^4+\omega_0^4)}{2b^{5/4}\pi\omega^4-c_\infty\pi\omega^2\omega_0\sqrt{b}[\sqrt{2}(a_2+a_0\sqrt{b})+a_1b^{1/4}]+\omega_0^3\left\{\pi c_\infty[\sqrt{2}(a_2-a_0\sqrt{b})+2b^{1/4}\omega_0]-a_1b^{1/4}c_\infty\ell n\left(b\frac{\omega^4}{\omega_0^4}\right)\right\}}. \tag{3}$$

Using the analytical representation for $c_{\text{phase}}(\omega)$ given in Eq. (3) in the least-squares data-fitting calculation, rather than carrying out the Hilbert transform of Eq. (2) numerically, provides a significant increase in computational efficiency, and thus computational speed.

B. Experiment

As a ''sanity check'' of the sound speed and attenuation measurement method, several sample panels fabricated from materials of known sound speed were first evaluated. (Only a small subset of these cases will be described here.) Good agreement with the known results was obtained. Acquiring multiple data sets for the case in which no panel was present provided a sensitive test for investigating the accuracy of the method. This test simulates the case of a sample with a perfect impedance match to the surrounding water medium and, since the speed of sound in fresh water is accurately known, the results for this case could be readily evaluated. In one typical test, the water temperature was measured to be 17.8 °C. According to an empirical formula,[7] the speed of sound in distilled water at this temperature is 1.4751×10^5 cm/s, which is taken here to be ''ground truth.''

Phase change and insertion loss data were acquired over the frequency interval 20 to 50 kHz in this experiment. Since no ''sample'' was actually present in these measurements, it was possible to assume various sample ''thicknesses'' in performing the least-squares fits. The results were as follows. For an assumed thickness of 1 cm, $c_\infty=1.476\,60\times10^5$ cm/s ± 199 cm/s, $\alpha=1.8\times10^{-3}$ nepers/cm $\pm 6.2\times10^{-3}$ nepers/cm. For an assumed thickness of 5 cm, $c_\infty=1.471\,90\times10^5$ cm/s ± 55 cm/s, $\alpha=5.0\times10^{-4}$ nepers/cm $\pm 1.5\times10^{-3}$ nepers/cm. And for an assumed thickness of 15 cm, $c_\infty=1.475\,20\times10^5$ cm/s ± 13 cm/s, $\alpha=1.0\times10^{-4}$ nepers/cm $\pm 4.5\times10^{-4}$ nepers/cm. The uncertainties were computed for 95% confidence limits, and were calculated using

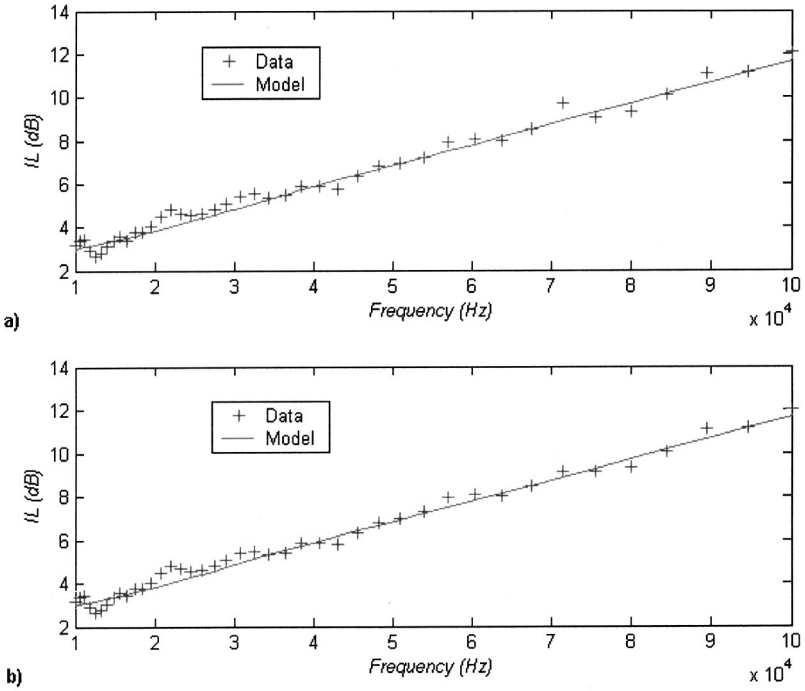


FIG. 5. Insertion loss for a dispersive sample. Solid lines are the results of the fit, plus (+) symbols denote data. (a) First measurement. (b) Repeat measurement.

standard methods associated with least-squares fitting.[8] Notice that the attenuation values (α) are all quite small, and are consistent with an assumption of zero attenuation (reasonable, for water), to within the computed uncertainties, since in all cases the uncertainties in the attenuation exceed its probable value. (The attenuation values given here represent the results computed for the a_0 constant of the model. The values obtained for a_1 and a_2 were sufficiently smaller than the given results that they did not significantly affect the attenuation over the frequency band considered in this test.) While the numerical values and uncertainty limits given for the c_∞ parameter for each of the assumed thicknesses are not entirely consistent, the inconsistencies are minor. (The reader should bear in mind that the parameter c_∞ denotes the phase speed at *infinite* frequency, while the measurements were only performed up to a frequency of 100 kHz.) Each of the measured values can be seen to be reasonable estimates of the sound speed obtained from the empirical formula for this test temperature.

With these successes obtained from evaluating known samples, the method was then applied to a more complex sample of unknown properties, one expected to be fairly lossy and dispersive. Panel thickness is 1.27 cm; panel density is 1.33 g/cm^3. The results of fitting the model to the data

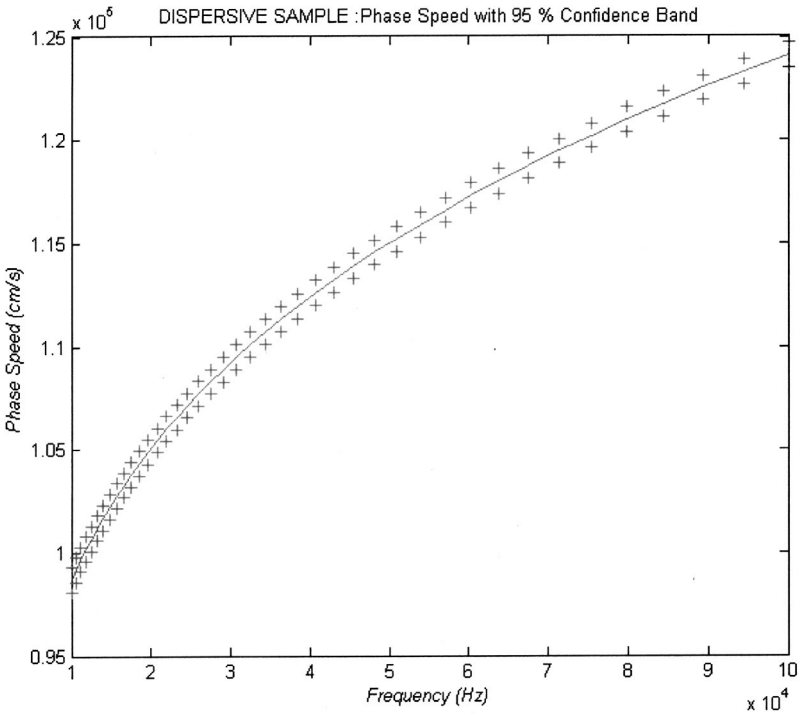


FIG. 6. Phase speed as a function of frequency deduced from the fits for the dispersive sample. Solid line is the result of substituting the four model parameters obtained from fitting the data [as illustrated in Figs. 4(a), (b) and 5(a), (b)] into Eq. (3). (The a_2 parameter was held fixed at a value of 10^{-5} nepers/cm in this fit.) The plus signs (+) denote the 95% confidence band, as determined from a standard propagation-of-error calculation for a least-squares procedure (Ref. 8).

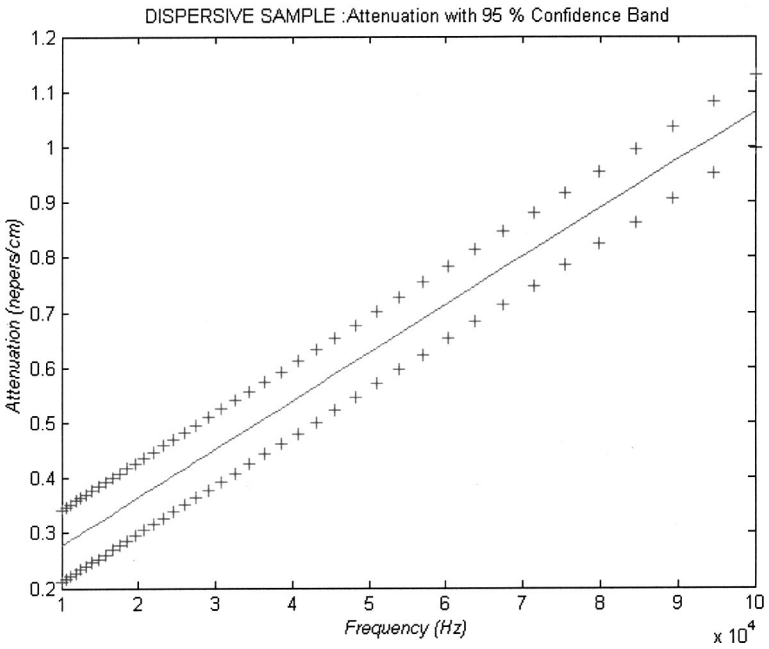


FIG. 7. Attenuation as a function of frequency deduced from the fits for the dispersive sample. Solid line is the result of substituting the four model parameters obtained from fitting the data [as illustrated in Figs. 4(a), (b) and 5(a), (b)] into Eq. (1). (The a_2 parameter was held fixed at a value of 10^{-5} nepers/cm in this fit.) The plus signs (+) denote the 95% confidence band, as determined from a standard propagation-of-error calculation for a least squares procedure (Ref. 8).

acquired from this sample are shown in Figs. 4 and 5. The model parameters obtained from the least-squares fitting process are as follows:

$c_\infty = 1.49151 \times 10^5$ cm/s $\pm$ 744 cm/s;
$a_0 = 0.1894$ nepers/cm $\pm$ 0.01159 nepers/cm;
$a_1 = 4.377$ nepers/cm $\pm$ 0.06556 nepers/cm; and
$a_2 = 10^{-5}$ nepers/cm $\pm$ 0 nepers/cm.

The quantity a_2 was held fixed at the indicated small numerical value, because an initial attempt to allow this parameter to vary in the least-squares calculation produced an unstable fit. (A negligibly small numerical value is used for an unvaried quantity like this, rather than a pure zero value, because zeros cause numerical difficulties in the software that implements the fitting calculations.) The value obtained for a_2 when it was permitted to vary was $a_2 = 1.375$ nepers/cm ± 2.599 nepers/cm. Since, as can be seen, the uncertainties are much greater than the probable value, it was felt that this indicated that the data were insensitive to this parameter, so it was held fixed in the manner noted above.

Substituting the four model parameters back into Eq. (1) and Eq. (3), and numerically evaluating these equations over the frequencies of the experiment, produced the results shown in Fig. 6 and Fig. 7 (solid lines.) The 95% confidence-limit uncertainty bands (denoted by plus signs in the figures), were again deduced using standard statistical techniques.[8]

V. SUMMARY AND CONCLUSION

A technique for measuring the phase change induced in an acoustic sound field when a panel is inserted between a projector and receiver was described. The method can determine the change in phase accurate to $\pm\frac{1}{2}$ deg of phase over the frequency interval of 10 to 100 kHz. A method for determining the phase speed and attenuation of a single-layer sample panel was also described. The method involves the adjustment of four or fewer parameters over the indicated measurement decade, and typically determines the phase speed accurate to ± 30 m/s. Extensions of the sound speed and attenuation measurement method to high hydrostatic pressure and multilayer panels are the subject of future work.

ACKNOWLEDGMENTS

We gratefully acknowledge helpful technical conversations with Stephen E. Forsythe and Robert M. Drake. This work was supported by the Office of Naval Research, Code 321.

[1] R. J. Bobber, *Underwater Electroacoustic Measurements* (U.S. Government Printing Office, Washington, 1970), pp. 287–299.
[2] Anonymous, *Underwater Electroacoustic Standard Transducers Catalogue*, USRD, Naval Research Laboratory (1982).
[3] R. P. Shaw and P. Bugl, "Transmission of plane waves through layered viscoelastic media," J. Acoust. Soc. Am. **46**, 649–654 (1968).
[4] C. W. Horton, "Dispersion relationships in sediments and sea water," J. Acoust. Soc. Am. **55**, pp. 547–549 (1974).
[5] J. C. Piquette, "Transmission coefficient measurement and improved sublayer material property determination for thick underwater acoustic panels: A generalization and improvement of the ONION method," J. Acoust. Soc. Am. **92**, 468–477 (1992).
[6] K. R. Waters, M. S. Hughes, J. Mobley, and J. G. Miller, "On the applicability of Kramers–Kronig relations for untrasonic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **108**, 556–563 (2000).
[7] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982), Eq. (5.22), p. 107.
[8] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in* FORTRAN (Cambridge University Press, New York, 1992), Chap. 15.

Highly directional acoustic receivers

Benjamin A. Cray[a] Victor M. Evora, and Albert H. Nuttall
Naval Undersea Warfare Center Division, 1176 Howell Street, Newport, Rhode Island 02891-1708

(Received 15 February 2002; revised 7 November 2002; accepted 16 December 2002)

The theoretical directivity of a single combined acoustic receiver, a device that can measure many quantities of an acoustic field at a collocated point, is presented here. The formulation is developed using a Taylor series expansion of acoustic pressure about the origin of a Cartesian coordinate system. For example, the quantities measured by a *second-order* combined receiver, denoted a dyadic sensor, are acoustic pressure, the three orthogonal components of acoustic particle velocity, and the nine spatial gradients of the velocity vector. The power series expansion, which can be of any order, is cast into an expression that defines the directivity of a single receiving element. It is shown that a single highly directional dyadic sensor can have a directivity index of up to 9.5 dB. However, there is a price to pay with highly directive sensors; these sensors can be significantly more sensitive to nonacoustic noise sources. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1543851]

PACS numbers: 43.58.Fm, 43.30.Xm, 43.20.Ye [SLE]

I. INTRODUCTION

Since the 1970s, multimicrophone probes have been routinely used to measure pressure and acoustic particle velocity in air. These probes measure sound intensity, which is a vector quantity defined as the average acoustic power transmitted per unit area in the direction of wave propagation. Typically, the probes were designed to measure a single component of the pressure gradient (such as $\Delta p/\Delta z$). This measurement was then used to estimate the acoustic velocity normal to the surface of a radiating structure. Acoustic intensity was obtained, as described by Chung[1] and Fahy,[2] by taking the imaginary part of the cross spectrum between the two collinear and closely spaced microphones. The technique required dual-channel spectrum analyzers, which were then commercially available. The intensity probes did not measure pressure and velocity directly at a collocated point—velocity was estimated using pressure gradient microphones. Thus, these sensors had, at some level, phase-mismatch errors and finite- difference approximation errors.

In-water measurements of the acoustic particle velocity vector and the associated intensity vector are less common, though sonobuoys have been used to measure pressure gradients in the horizontal plane for many years. Over the past decade, Lo,[3] Ng,[4] Shchurov,[5] D'Spain,[6] Wei,[7] and other researchers[8–13] have either built underwater vector sensors or have analyzed and measured such quantities as acoustic intensity, acoustic energy density and flux, and acoustic particle motion polarization. Algorithms that can be used to beamform an array of vector sensors, for detection and for direction of arrival estimation, are described by D'Spain,[14] Nehorai,[15] and Cray.[16]

The goal of the work presented here is to determine the directivity of various types of combined receivers, that is, sensors that measure acoustic pressure as well as additional components of the acoustic field (such as particle velocity and velocity gradients). For pressure-sensing hydrophones, directivity is fundamentally limited by the size of the aperture of an array of hydrophones; therefore, increasing the aperture increases directivity. Combined receivers, however, provide additional directivity gain by measuring additional quantities of the acoustic field.

II. TAYLOR SERIES EXPANSION OF AN ACOUSTIC PRESSURE FIELD

An nth order Taylor series[17] expansion of an acoustic pressure field contained within a finite region (h,k,l) about the point (x_o,y_o,z_o) can be expressed as

$$p(x_o+h,\ y_o+k,z_o+l) = p(x_o,y_o,z_o)+\sum_{n=1}^{\infty}\frac{1}{n!}\left[\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}+l\frac{\partial}{\partial z}\right)^n \times p(x,y,z)\right]_{x=x_o,y=y_o,z=z_o}. \quad (1)$$

Here, the pressure field's time-harmonic dependence, $\exp(i\omega t)$, is assumed and suppressed. Expanding the series to second order ($n=2$) and letting $x=x_o+h$, $y=y_o+k$, and $z=z_o+l$ yields

$$p(\mathbf{r})\approx p(\mathbf{r}_o)+[\mathbf{r}-\mathbf{r}_o]\begin{bmatrix}\frac{\partial p}{\partial x}\\ \frac{\partial p}{\partial y}\\ \frac{\partial p}{\partial z}\end{bmatrix} + \frac{1}{2}[\mathbf{r}-\mathbf{r}_o]\begin{bmatrix}\frac{\partial^2 p}{\partial x^2} & \frac{\partial}{\partial y}\frac{\partial p}{\partial x} & \frac{\partial}{\partial z}\frac{\partial p}{\partial x}\\ \frac{\partial}{\partial x}\frac{\partial p}{\partial y} & \frac{\partial^2 p}{\partial y^2} & \frac{\partial}{\partial z}\frac{\partial p}{\partial y}\\ \frac{\partial}{\partial x}\frac{\partial p}{\partial z} & \frac{\partial}{\partial y}\frac{\partial p}{\partial z} & \frac{\partial^2 p}{\partial z^2}\end{bmatrix}[\mathbf{r}-\mathbf{r}_o]^T, \quad (2)$$

where row vectors $\mathbf{r}$ and $\mathbf{r}_o$ are defined, respectively, as $\mathbf{r}$

[a]Electronic mail: crayba@npt.nuwc.navy.mil

$=[xyz]$ and $\mathbf{r}_o=[x_o y_o z_o]$; and matrix notation has been used to simplify the series expression.

The spatial gradients of pressure, shown in the expansion's first-order term, may be related to acoustic particle velocity via the linearized momentum equations, i.e.,

$$\frac{\partial p}{\partial x} = -\rho_o i\omega u, \tag{3a}$$

$$\frac{\partial p}{\partial y} = -\rho_o i\omega v, \tag{3b}$$

and

$$\frac{\partial p}{\partial z} = -\rho_o i\omega w, \tag{3c}$$

where ρ_o is the ambient density of the surrounding fluid, ω is the radian frequency of the acoustic plane wave, and u, v, and w are the orthogonal components of the acoustic particle velocity vector. By substituting Eqs. (3a) through (3c) into Eq. (2), and taking the appropriate partial derivatives, Eq. (2) can be rewritten as

$$p(\mathbf{r}) \approx p(\mathbf{r}_o) + \rho_o i\omega[\mathbf{r}_o - \mathbf{r}]\begin{bmatrix} u \\ v \\ w \end{bmatrix} + \frac{1}{2}\rho_o i\omega[\mathbf{r}_o - \mathbf{r}]\begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} & \frac{\partial u}{\partial z} \\ \frac{\partial u}{\partial y} & \frac{\partial v}{\partial y} & \frac{\partial v}{\partial z} \\ \frac{\partial u}{\partial z} & \frac{\partial v}{\partial z} & \frac{\partial w}{\partial z} \end{bmatrix}[\mathbf{r} - \mathbf{r}_o]^T. \tag{4}$$

Notice that the second-order off-diagonal terms above have been simplified and are symmetric. The propagating plane wave is irrotational and, hence, the curl of the velocity vector is zero ($\partial u/\partial y = \partial v/\partial x$, $\partial u/\partial z = \partial w/\partial x$, $\partial v/\partial z = \partial w/\partial y$).

It is evident that each term in the series is proportional to different characteristics of the acoustic field. The zeroth-order term is proportional to acoustic pressure, the quantity that is commonly measured by hydrophones and acoustic arrays. The first-order term provides information on the three components of acoustic particle velocity. The trace of the second-order term can be shown to be proportional to the instantaneous density of the acoustic field from the linearized continuity equation, along with the typical acoustic approximation for condensation ($s=(\rho-\rho_o)/\rho_o \ll 1$); thus

$$\rho(x,y,z;t) = \frac{i\rho_o}{\omega}\left\{\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right\}. \tag{5}$$

Clearly, each term in the series provides additional information on the acoustic field, and a combined receiver will measure, at a collocated point (x_o, y_o, z_o), these additional quantities of the acoustic field. Thus, the series expansion illustrates that a single receiver can be used to determine properties of the acoustic field, not simply at a point, but within a volumetric region surrounding the sensor. Many investigators[5,6,18–22] have built and verified the performance of vector sensors, or combined receivers, that measure acoustic pressure and the three orthogonal components of acoustic particle velocity. However, measuring the quantities shown in the higher-order terms will be more challenging; the relative magnitudes of these quantities are much less than the magnitude of acoustic pressure given in the zeroth-order term. Consider, for example, a simplified Taylor series to the third order along the x axis:

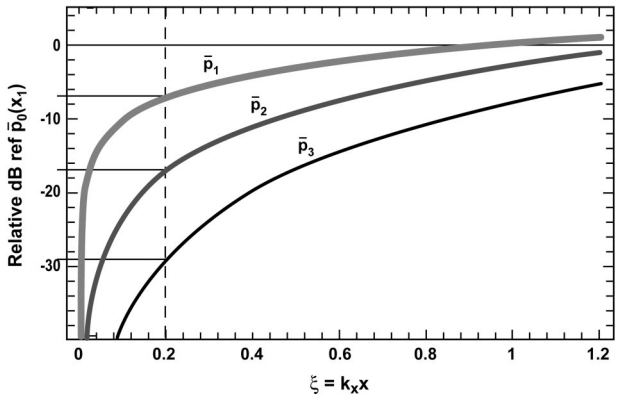


FIG. 1. Comparison of the relative rms pressures of correction terms in a simplified Taylor series.

$$p(x,t) = p(0,t) + x\left.\frac{\partial p}{\partial x}\right|_{x=0} + \frac{x^2}{2}\left.\frac{\partial^2 p}{\partial x^2}\right|_{x=0} + \frac{x^3}{6}\left.\frac{\partial^3 p}{\partial x^3}\right|_{x=0}. \tag{6}$$

Assuming harmonic plane waves of the form $p(x,t) = P_o \sin(\omega t - k_x x + \phi)$ with $k_x = k_o \cos(\theta)$, the expansion can be rewritten as

$$p(x,t) \cong P_o\left\{\sin(\omega t+\phi) - \xi\cos(\omega t+\phi) - \frac{\xi^2}{2}\sin(\omega t+\phi) - \frac{\xi^3}{6}\cos(\omega t+\phi)\right\}, \tag{7}$$

where, for $\theta = 0$, $\xi = k_o x = 2\pi(x/\lambda)$.

The quantity ξ determines the error bounds associated with the expansion. For a given acceptable error limit (say, $\xi=0.2$), the region of a valid series approximation increases with decreasing frequency, or, conversely, increasing acoustic wavelength.

The rms pressure of each term is

$$\overline{p_0}(x) = \sqrt{\frac{P_o}{2}}, \quad p_1(x) = \sqrt{\frac{P_o}{2}}\xi, \quad \overline{p_2}(x) = \sqrt{\frac{P_o}{2}}\frac{\xi^2}{2}, \quad p_3(x) = \sqrt{\frac{P_o}{2}}\frac{\xi^3}{6}. \tag{8}$$

Figure 1 compares the relative rms pressures of each expansion term. As shown, for $\xi=0.2$, the rms pressure of the first-order correction term, $p_1(x)$, is approximately 7 dB less than that of the zeroth-order direct-pressure measurement. The rms pressure of the second- and third-order terms are, respectively, 17 and 29 dB below that of the conventional measurement of pressure. This result indicates that the

magnitudes of the higher-order correction terms are small compared to the zeroth-order term and diminish rapidly with increasing order. The rms pressure within a given Taylor series approximation region is primarily determined by the zeroth-order term. Hence, if the goal is to approximate the amplitude of acoustic pressure surrounding a point, it is sufficient to limit the measurement to zeroth order. As is, the Taylor series expansion does not illustrate the advantage of the directional characteristics of particle velocity and the higher-order velocity gradients.

In addition, each of the higher-order quantities has their own self-noise characteristics and are more sensitive to subsonic noise sources.[25–29]

III. DIRECTIVITY FACTORS OF VARIOUS COMBINED RECEIVERS

The directional dependence of the acoustic particle velocity components, as well as the velocity gradient components, may be obtained from the linearized momentum and continuity equations. Thus, assuming a standard coordinate system:

$$p(x,y,z;t)=p_o\exp\{ik_o[x\cos\theta\sin\phi+y\sin\theta\sin\phi+z\cos(\phi)]+i\omega t\}, \tag{9}$$

$$u(x,y,z;t)=\frac{-\cos(\theta)\sin(\phi)}{\rho_o c}p(x,y,z;t), \tag{10}$$

$$v(x,y,z;t)=\frac{-\sin(\theta)\sin(\phi)}{\rho_o c}p(x,y,z;t), \tag{11}$$

and

$$w(x,y,z;t)=\frac{-\cos(\phi)}{\rho_o c}p(x,y,z;t), \tag{12}$$

where c is the sound speed in the fluid and p is pressure. Similarly, the gradients of Eqs. (10), (11), and (12) are

$$u'(x,y,z;t)=\frac{\partial u}{\partial x}=\frac{-i\omega}{\rho_o c^2}\cos^2(\theta)\sin^2(\phi)p(x,y,z;t), \tag{13}$$

$$v'(x,y,z;t)=\frac{\partial v}{\partial y}=\frac{-i\omega}{\rho_o c^2}\sin^2(\theta)\sin^2(\phi)p(x,y,z;t), \tag{14}$$

$$w'(x,y,z;t)=\frac{\partial w}{\partial z}=\frac{-i\omega}{\rho_o c^2}\cos^2(\phi)p(x,y,z;t). \tag{15}$$

A scale factor, denoted q_1 and set equal to the negative of the fluid's specific acoustic impedance ($\rho_o c$), is now introduced to convert Eqs. (10), (11), and (12) into *equivalent* pressure amplitudes. Thus

$$q_1 u=\cos(\theta)\sin(\phi)p(x,y,z;t), \tag{16}$$

$$q_1 v=\sin(\theta)\sin(\phi)p(x,y,z;t), \tag{17}$$

$$q_1 w=\cos(\phi)p(x,y,z;t). \tag{18}$$

The scaled velocity components can now be weighted and summed along with the acoustic pressure. The power sum is then defined as

$$B^{(4)}(\theta,\phi)=|w_p+w_x\cos(\theta)\sin(\phi)+w_y\sin(\theta)\sin(\phi)+w_z\cos(\phi)|^2, \tag{19}$$

where the arbitrary weighting coefficients $\{w_p,w_x,w_y,w_z\}$ include the scaling factor q_1, and p_o has been set to unity.

In a similar manner, the second-order terms may be summed with the quantities in Eq. (19). However, the scaling factor is more involved, being both complex ($\pi/2$ phase shift) and frequency dependent. The scaling factor is defined as $q_2=i\rho_o c^2/\omega$. First consider the diagonal terms of the Hessian matrix

$$q_2\left(\frac{\partial u}{\partial x}\right)=\cos^2(\theta)\sin^2(\phi)p(x,y,z;t), \tag{20}$$

$$q_2\left(\frac{\partial v}{\partial y}\right)=\sin^2(\theta)\sin^2(\phi)p(x,y,z;t), \tag{21}$$

and

$$q_2\left(\frac{\partial w}{\partial z}\right)=\cos^2(\phi)p(x,y,z;t). \tag{22}$$

The power sum of the seven weighted quantities can now be written as

$$B^{(7)}(\theta,\phi)=|w_p+w_x a+w_y b+w_z c+w_{xx}a^2+w_{yy}b^2+w_{zz}c^2|^2, \tag{23}$$

where $a=a(\theta,\phi)=\cos(\theta)\sin(\phi)$, $b=b(\theta,\phi)=\sin(\theta)\sin(\phi)$, and $c=c(\phi)=\cos(\phi)$. Similarly, the power sum of the complete ten-term weighted expansion of the Hessian matrix is

$$B^{(10)}(\theta,\phi)=|w_p+w_x a+w_y b+w_z c+w_{xx}a^2+w_{yy}b^2+w_{zz}c^2+w_{xy}2ab+w_{xz}2ac+w_{yz}2bc|^2. \tag{24}$$

The optimal weights for a vector sensor have been previously derived[16] by solving for the set of weights that produced an extremum value for directivity factor. Theoretically, the direct elimination procedure can be extended to determine the optimal weights for higher-order sensors, that is, the weights w_{xx}, w_{yy}, w_{zz}, w_{xy}, w_{xz}, w_{yz}. However, the algebra becomes cumbersome, even with the aid of programs such as MATHEMATICA™. Instead, the optimal weights for a simplified directional sensor that measures pressure, the three components of particle velocity, and a *single* gradient of velocity (i.e., the weights w_p, w_x, w_y, w_z, w_{xx}) were determined here. These optimal weights generate the directivity shown in Fig. 2, which is noted to be as much as 9.5 dB (or 3.5 dB greater than that of a vector sensor) along the boresight of the x axis. Also shown are the steering directions at which the velocity gradient $\partial u/\partial x$ did not provide additional gain. Not surprisingly, if the gradient $\partial v/\partial y$ is measured as well (allowing for the additional weight w_{yy}), the directivity increases along the boresight of the y axis, at $\theta_s=\pm 90°$, as shown in Fig. 3. The beam response of this six-component directional sensor, using optimal weighting, is illustrated in Fig. 4. The half-power beamwidth is 65°. For comparison, Fig. 5 shows the optimal response for a four-component vector sensor.

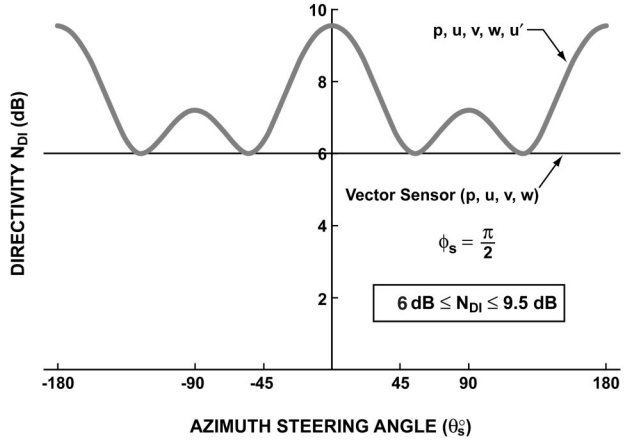


FIG. 2. Directivity of a five-component receiver.

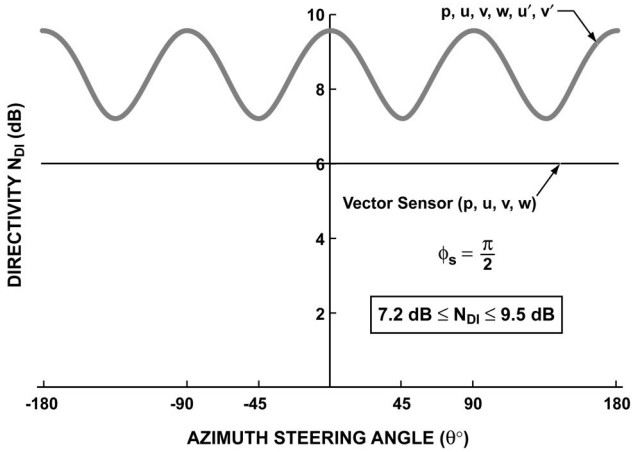


FIG. 3. Directivity of a six-component receiver.

The weight set for optimal directivity of a vector sensor [Eq. (19)] is given as $\{w_p=1,\ w_x=3a_s,\ w_y=3b_s,\ w_z=3c_s\}$, where $a_s=a(\theta_s,\phi_s)$, $b_s=b(\theta_s,\phi_s)$, and $c_s=c(\phi_s)$, and θ_s and ϕ_s, respectively, are the chosen azimuth and elevation steering, or look, directions. For the five-component receiver, the optimal weight set is $\{w_p=1,\ w_x=4a_s/(3-5a_s^2),\ w_y=4b_s/(3-5a_s^2),\ w_z=4c_s/(3-5a_s^2),\ w_{xx}=5(3a_s^2-1)/(3-5a_s^2)\}$; and for the six-component receiver: $\{w_p=1,\ w_x=2a_s/(4-5\sin^2(\phi_s)),\ w_y=2b_s/(4-5\sin^2(\phi_s)),\ w_z=2c_s/(4-5\sin^2(\phi_s)),\ w_{xx}=5(2a_s^2+b_s^2-1)/(4-5\sin^2(\phi_s)),\ w_{yy}=5(2a_s^2+a_s^2-1)/(4-5\sin^2(\phi_s))\}$.

The optimal weight set of the four-component vector sensor reduced to a constant (i.e., 3) times the directional response of each velocity component. Given this, a similar weighting set may be considered[30] for a ten-component directional receiver, now allowing for two constants μ_1 and μ_2. That is, the optimal weight set is assumed to be of the form $\{w_p=1,\ w_x=\mu_1 a_s,\ w_y=\mu_1 b_s,\ w_z=\mu_1 c_s,\ w_{xx}=\mu_2 a_s^2,\ w_{yy}=\mu_2 b_s^2,\ w_{zz}=\mu_2 c_s^2,\ w_{xy}=2\mu_2 a_s b_s,\ w_{xz}=2\mu_2 a_s c_s,\ w_{yz}=2\mu_2 b_s c_s\}$. With this assumption, the algebra required to determine the optimal weights is simplified. As defined in Ref. 16, the directivity factor, DF, of a combined receiver is

$$\mathrm{DF}=\frac{4\pi B(\theta_s,\phi_s)}{\int_0^{2\pi}\int_0^{\pi}B(\theta,\phi)\sin(\phi)\,d\phi\,d\theta}. \tag{25}$$

Substituting the ten-component receiver response, Eq. (24), into Eq. (25) and performing the integration (with $w_p=1$) yields

$$\mathrm{DF}^{(10)}=\frac{15B^{10}(\theta_s,\phi_s)}{[15+5w_x^2+3w_{xx}^2+4w_{xy}^2+4w_{xz}^2+5w_y^2+3w_{yy}^2+4w_{yz}^2+5w_z^2+2w_{yy}(5+w_{zz})+2w_{xx}(5+w_{yy}+w_{zz})+w_{zz}(10+3w_{zz})]}. \tag{26}$$

The optimal weight set could be found by maximizing the above ratio for ten arbitrary weights; the procedure would lead to solving ten equations with ten unknowns. However, with the assumed optimal weight set for the dyadic sensor, the directivity factor DF reduces to

$$\mathrm{DF}^{(10)}=\frac{15(1+\mu_1+\mu_2)^2}{15+(5\mu_1+\mu_2(10+3\mu_2)}, \tag{27}$$

which is independent of steering angle.

It is conjectured that maximum directivity is 9.5 dB, which occurs when DF=9. Hence, setting $\mathrm{DF}^{(10)}=9$, and solving for μ_1 in terms of μ_2,

$$\mu_1=\tfrac{1}{10}\{5+5\mu_2\pm\sqrt{15}\sqrt{-(\mu_2+5)^2}\}. \tag{28}$$

A real solution for constants μ_1 and μ_2 is obtained only for $\mu_2=-5$; then, the two constants for the optimal weights are ($\mu_1=-2$, $\mu_2=-5$).

Optimal weights for higher-order receivers can be obtained (see the Appendix) numerically by forming the Lagrangian that minimizes the denominator of Eq. (25), thus maximizing the directivity factor. In principle, this numerical procedure will determine the directivity of any arbitrary order combined receiver.

IV. CONCLUSIONS

Table I compares the directivity indices of various types of acoustic receivers. A single pressure sensor is omnidirectional, providing no spatial discrimination of an acoustic field. Single-axis velocity sensors, measuring the component velocities u, v, w are dipoles, or cosine-directive. In the free field, these velocity sensors have a directivity index (N_{DI}) of 4.8 dB and can provide spatial filtering of an acoustic field. Without a corresponding measurement of pressure, however, velocity sensors have directional ambiguity—they

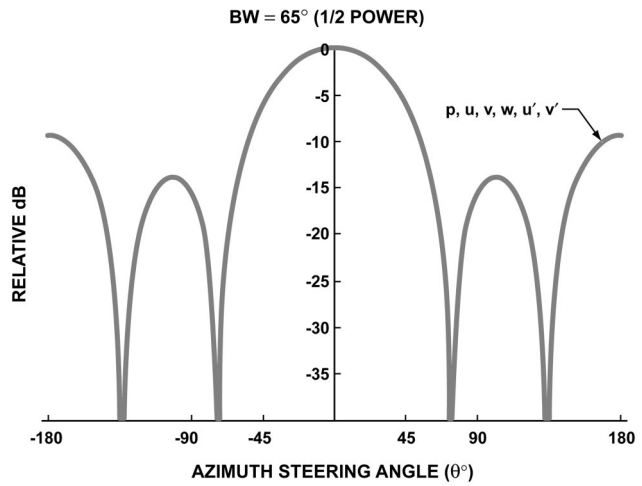


FIG. 4. Optimal combined receiver (six-component) beam response.

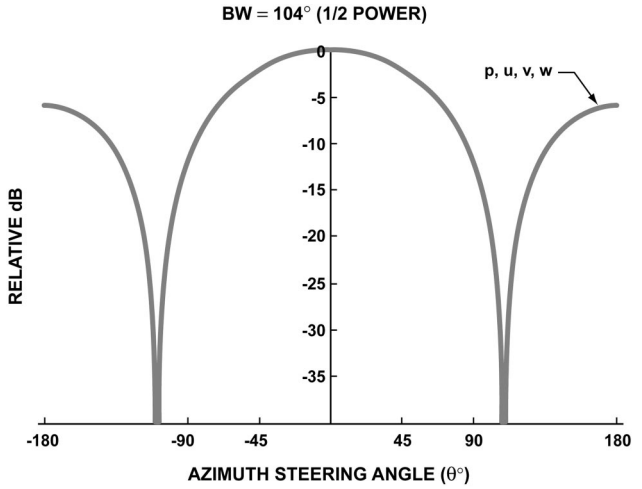


FIG. 5. Optimal vector sensor beam response (N_{DI}=6 dB).

cannot determine ''left/right'' propagation. True pressure-gradient sensors (referred to as $p-p$ probes[9]) provide a first-order finite-difference estimate of particle velocity and, given that these sensors measure pressure directly, can determine arrival direction.

Vector sensors, which measure pressure and the three components of acoustic particle velocity (**p**, **u**, **v**, **w**), have a directivity index of 6 dB. The optimal beam response of a vector sensor was illustrated in Fig. 5, with half-power mainlobe beamwidth of 104°. Higher-order sensors measure gradients of acoustic particle velocity; it may be shown that these gradients (for acoustic plane waves) are proportional to variations in the instantaneous density of the acoustic field. Just as pressure-gradient sensors infer acoustic particle velocity, velocity-gradient sensors appear to infer changes in density. If it were possible to measure density variations directly (as, for example, accelerometers measure acoustic particle acceleration directly), then it would not be necessary to form finite-difference approximations of these gradients with lower-order components.

A five-component sensor (**p**, **u**, **v**, **w**, **u**′, where $\mathbf{u}'=\mathbf{u}_x$) can have a directivity index of 9.5 dB, which is nine times the DF of an omnidirectional pressure sensor. However, this directivity depends on the arrival direction of the acoustic plane wave. Maximum directivity (9.5 dB) is obtained only along the boresight of the x axis. As additional velocity gradients are measured, the angular dependence is eliminated and a constant maximum directivity of 9.5 dB is obtained.

Combined receivers measure additional quantities of the acoustic field; hence, the number of processing channels increases with the order of the receiver. Combined receivers that measure pressure and each orthogonal component of acoustic particle acceleration have been built by many researchers, including Shchurov,[5] EDO Corporation,[19] Wilcoxon Research,[20] the University of Massachusetts-Dartmouth,[21] AcousTech Corporation,[22] and Pennsylvania State University.[23,24]

McConnell[31] has designed and tested an underwater acoustic intensity sensor, denoted as the **u**–**u** probe, that estimates the velocity gradient ($\partial w/\partial z$) by using the finite difference approximation $\Delta w/\Delta z$. The measurement is analogous to the standard pressure gradient technique. Measurements made of a reactive intensity field with the **u**–**u** probe compared well with those made from a **p**–**u** probe, and it was concluded that the **u**–**u** technique was viable. Thus, it may be possible in the future to measure the second-order expansion terms shown in Eq. (4), and, hence, increase an array's directivity without increasing the array's acoustic aperture.

It will remain a challenge, though, to measure these higher-order components. Additional self-noise mechanisms, not present with the lower-order acoustic components, can corrupt the acoustic sensitivity of highly directive receivers. The gains possible with these sensors are significant. The point directional dyadic sensors described here have directivity gains of as much as 9.5 dB. In addition, multiple nulls may be generated within a single directional sensor's angular response. These nulls can then be used to reduce signal interference, acoustic and nonacoustic sensor self noise, and unwanted multipath reflections.

ACKNOWLEDGMENTS

The authors would like to thank the Office of Naval Research, Program Officer Dr. Roy Elswick (ONR 321S), for his financial support of this research. In addition, the authors

TABLE I. Free-field directivity indices for various acoustic receivers.

Combined receivers:	p	u,v,w	p,u,v,w	p,u,v,w,u'	p,u,v,w,u',v',w'	$p,u,v,w,u',v',w',u_{xy},u_{xz},v_{xz}$
Directivity index (N_{DI})	0 dB	4.8 dB	6 dB	6 dB$\leqslant N_{\mathrm{DI}}\leqslant$9.5 dB	7.2 dB$\leqslant N_{\mathrm{DI}}\leqslant$9.5 dB	9.5 dB
No. of channels	1	3	4	5	7	10

appreciate the detailed comments made by each reviewer of this manuscript and are grateful for the pertinent references they identified.

APPENDIX: OPTIMAL COMPONENT WEIGHTS VIA LAGRANGE MULTIPLIERS

Consider a limited combined receiver that measures pressure gradients, to any order, only along the z axis. Thus, for a harmonic plane wave

$$\frac{\partial^n p}{\partial z^n} = (ik_o c(\phi))^n p(x,y,z,t), \tag{A1}$$

where $p(x,y,z,t) = P_o \exp\{ik_o((xa(\theta,\phi) + yb(\theta,\phi), +zc(\phi)) + iwt)\}$.

The combined receiver is located at the origin of the coordinate system, and, assuming a unit-amplitude pressure, $P_o = 1$, the power sum of the weighted and scaled gradient may be written as

$$B(\theta,\phi) = \left| \sum_{n=0}^{N} w_n c^n(\phi) \right|^2 = \sum_{n=0}^{N} \sum_{m=0}^{N} w_n w_m c^n(\phi) c^m(\phi), \tag{A2}$$

with the harmonic-time dependence omitted.

The directivity factor (DF) thus becomes

$$\mathrm{DF} = \frac{4\pi B(\theta_s,\phi_s)}{\int_0^{2\pi}\int_0^{\pi} \Sigma_{n=0}^{N} \Sigma_{m=0}^{N} w_n w_m c^n(\phi) c^m(\phi) \sin(\phi)\, d\phi\, d\theta}, \tag{A3}$$

where

$$B(\theta_s,\phi_s) = \left| \sum_{n=0}^{N} w_n \cos^n(\phi_s) \right|^2$$

and (θ_s,ϕ_s) defines the arrival angles of the signal.

Interchanging the order of summation and integrating over angles θ, ϕ yields

$$\mathrm{DF} = \frac{(\Sigma_{n=0}^{N} \cos^n(\phi_s) w_n)^2}{\sum_{n=0}^{N} \sum_{m=0}^{N} \frac{w_n w_m}{2} \frac{(1-(-1)^{n+m+1})}{(n+m+1)}} = \frac{(\Sigma_{n=0}^{N} \alpha_n w_n)^2}{\Sigma_{n=0}^{N} \Sigma_{m=0}^{N} \beta_{nm} w_n w_m} \tag{A4}$$

with $\alpha_n = \cos^n(\phi_s)$ and $\beta_{nm} = [1-(-1)^{n+m+1}]/[2(n+m+1)]$.

The objective now is to maximize Eq. (A4) by choice of the weight set $\{w_n\}$. To simplify, introduce matrix notation $\mathbf{A} = [\alpha_o, \alpha_1, \ldots, \alpha_N]^T$, $\mathbf{W} = [w_o, w_1, \ldots, w_N]^T$, and $B = [\beta_{nm}]$, which is symmetric and positive definite, as may be determined from the denominator of (A3). Thus

$$\mathbf{DF} = \frac{(\mathbf{A^T W})^2}{\mathbf{W^T B W}}. \tag{A5}$$

From the Lagrangian, which minimizes $\mathbf{W^T B W}$, with the constraint $\mathbf{A^T W} = 1$, and thus maximizes DF:

$$\mathbf{q} = \mathbf{W^T B W} - \boldsymbol{\lambda} \mathbf{A^T W}. \tag{A6}$$

Equivalently, the Lagrangian may be expressed as

$$\mathbf{q} = \left(\mathbf{W} - \frac{\boldsymbol{\lambda}}{2} \mathbf{B^{-1} A} \right)^{\mathbf{T}} \mathbf{B} \left(\mathbf{W} - \frac{\boldsymbol{\lambda}}{2} \mathbf{B^{-1} A} \right) - \frac{\boldsymbol{\lambda}^2}{4} \mathbf{A^T B^{-1} A}. \tag{A7}$$

By inspection, the optimal weights are clearly $\mathbf{W_o} = (\boldsymbol{\lambda}/2)\, \mathbf{B^{-1} A}$, which yields a minimum value $\mathbf{q} = -(\boldsymbol{\lambda}^2/4)\, \mathbf{A^T B^{-1} A}$. From the constraint equation $(\mathbf{A^T W_o} = 1)$, the Lagrange multiplier is obtained

$$\frac{\boldsymbol{\lambda}}{2} = \frac{1}{\mathbf{A^T B^{-1} A}}, \tag{A8}$$

and the weights that generate optimal directivity for an arbitrary-order combined receiver are

$$\mathbf{W_o} = \frac{\mathbf{B^{-1} A}}{\mathbf{A^T B^{-1} A}}, \tag{A9}$$

and the corresponding maximum directivity (with $\mathbf{W_o}$ above) becomes

$$\mathbf{DF_o} = \mathbf{A^T B^{-1} A}. \tag{A10}$$

[1] J. Y. Chung, "Cross-spectral method of measuring acoustic intensity without errors caused by instrument phase mismatch," J. Acoust. Soc. Am. **64**, 1613–1616 (1978).

[2] F. J. Fahy, "Measurement of acoustic intensity using the cross-spectral density of two microphone signals," J. Acoust. Soc. Am. **62**, 1057–1059 (1977).

[3] E. Y. Lo and M. C. Junger, "Signal to Noise Ratio Enhancement by Underwater Intensity Measurement," J. Acoust. Soc. Am. **82**, 1450–1454 (1987).

[4] K. W. Ng, "Sound-intensity probe for underwater acoustic measurements," Proc. NOISE-CON 91, Institute of Noise Control Engineers, 1991, pp. 591–596 (1991).

[5] V. A. Shchurov and M. V. Kuyanova, "Use of acoustic intensity measurements in underwater acoustics (modern state and prospects)," Chin. J. Acoust. **18**(4), 315–326 (1999).

[6] G. L. D'Spain, W. S. Hodgkiss, and G. L. Edmonds, "The simultaneous measurement of infrasonic acoustic particle velocity and acoustic pressure in the ocean by freely drifting swallow floats," IEEE J. Ocean. Eng. **16**(2), 195–207 (1991).

[7] W. Wei, "Underwater measurement of the sound-intensity vector: Its use in locating sound sources, and in measuring the sound power of stationary and living sources," Ph.D. dissertation, University of Mississippi (UMI Dissertation Service Order Number 9431588) (1994).

[8] K. T. Wong and D. Zoltowski, "Closed-form underwater acoustic direction- finding with arbitrarily spaced vector hydrophones at unknown locations," IEEE J. Ocean. Eng. **22**(3), 566–575 (1997).

[9] 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).

[10] B. A. Cray and M. T. Silvia, "Conventional and adaptive beamforming with acoustic vector sensors," in Proceedings of the Undersea Defense Technology (UDT) Pacific 2000 Conference, February 2000, Sydney, Australia.

[11] M. A. Josserand and C. Maerfeld, "PVF2 velocity hydrophones," J. Acoust. Soc. Am. **78**, 861–867 (1985).

[12] 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).

[13] "Acoustic Particle Velocity Sensors: Design, Performance, and Applications," edited by M. Berliner and J. Lindberg, American Institute of Physics Conference Proceedings 368 (1995).

[14] G. L. D'Spain, W. S. Hodgkiss, G. L. Edmonds, J. C. Nickles, F. H. Fisher, and R. A. Harriss, "Initial analysis of the data from the vertical

DIFAR array,'' in Proceedings of Mastering the Oceans Through Technology (OCEANS 92), Newport, RI (1992).
[15] A. Nehorai and E. Paldi, ''Acoustic vector-sensor array processing,'' IEEE Trans. Signal Process. **41**(9), 2481–2491 (1994).
[16] B. A. Cray and A. H. Nuttall, ''Directivity factors for linear arrays of velocity sensors,'' J. Acoust. Soc. Am. **110**, 324–331 (2001).
[17] A. Taylor and R. Mann, *Advanced Calculus*, 3rd ed. (Wiley, New York, 1983).
[18] D. Hutt, P. Hines, and A. Hamilton, ''Measurements of underwater sound intensity,'' in *Proceedings of Oceans '99* (MTS/IEEE, Seattle), pp. 717–722 (1999).
[19] P. D. Baird, ''EDO directional acoustic sensor technology,'' on the CD-ROM *Proceedings of the Workshop on Directional Acoustic Sensors* (Naval Undersea Warfare Center Division, Newport, RI, 2001).
[20] P. A. Wlodkowski and F. Schloss, Advances in acoustic particle velocity sensorics,'' on the CD-ROM *Proceedings of the Workshop on Directional Acoustic Sensors* (Naval Undersea Warfare Center Division, Newport, RI, 2001).
[21] B. Aronov and D. Brown, ''An investigation of pressure gradient fixed and motion type transducers for underwater acoustics,'' Joint Technical Report BT-2001- 04-2, BTECH Acoustics, East Providence, RI, and University of Massachusetts, Dartmouth, MA (2001).
[22] J. A. McConnell, ''Practical experiences with inertial type underwater acoustic intensity probes,'' in Proceedings of the 2002 International Congress and Exposition on Noise Control Engineering (Inter-Noise 2002), Dearborn, MI, 19–22 August 2002 (ISBN Ino2).
[23] K. G. Kim, G. C. Lauchle, and T. B. Gabrielson, ''A pressure-acceleration underwater acoustic intensity sensor,'' Pennsylvania State University, Applied Research Laboratory Technical Report, TM 99-065 (May 1999).
[24] K. Kim, G. C. Lauchle, and T. B. Gabrielson, ''Pressure-particle acceleration underwater acoustic intensity sensor,'' J. Acoust. Soc. Am. **106**, 2297(A) (1999).
[25] B. A. Cray, ''Directional Acoustic Receivers: Signal and Noise Characteristics,'' on the CD-ROM *Proceedings of the Workshop on Directional Acoustic Sensors* (Naval Undersea Warfare Center Division, Newport, RI, 2001).
[26] G. Mardonik and W. J. Becker, ''Primitive comparison of the signal-to-noise ratios of pressure and velocity planar arrays,'' Report NSWCCD-516-97/256-7030, Naval Surface Warfare Center, Carderock Division (1997).
[27] B. D. Keller, ''Gradient hydrophone flow noise,'' J. Acoust. Soc. Am. **62**, 205–208 (1977).
[28] R. A. Finger, L. A. Abbagnaro, and B. B. Bauer, ''Measurements of low-velocity flow noise on pressure and pressure gradient hydrophones,'' J. Acoust. Soc. Am. **65**, 1407–1412 (1979).
[29] J. F. McEachern and G. C. Lauchle, ''Flow-induced noise on a bluff body,'' J. Acoust. Soc. Am. **97**, 947–953 (1995).
[30] M. T. Silvia, R. E. Franklin, and D. J. Schmidlin, ''Signal processing considerations for a general class of directional acoustic sensors,'' on the CD-ROM *Proceedings of the Workshop on Directional Acoustic Sensors* (Naval Undersea Warfare Center Division, Newport, RI, 2001).
[31] J. A. McConnell, G. C. Lauchle, and T. B. Gabrielson, ''Two geophone underwater acoustic intensity probe,'' U.S. Patent No. 6,172,940, 9 January 2001.

Sound focusing in rooms: The time-reversal approach

Sylvain Yon, Mickael Tanter,[a] and Mathias Fink
Laboratoire Ondes et Acoustique, Université Paris VII/ESPCI, CNRS UMR 7587, 10 rue Vauquelin, 75231 Paris Cedex 05, France

(Received 29 July 2002; accepted for publication 29 November 2002)

New perspectives in audible range acoustics, such as virtual sound space creation and active noise control, rely on the ability of the rendering system to recreate precisely a desired sound field. This ability to control sound in a given volume of a room is directly linked to the capacity to focus acoustical energy both in space and time. However, sound focusing in rooms remains a complicated problem, essentially because of the multiple reflections on obstacles and walls occurring during propagation. In this paper, the technique of time-reversal focusing, well known in ultrasound, is experimentally applied to audible range acoustics. Compared to classical focusing techniques such as delay law focusing, time reversal appears to considerably improve quality of both temporal and spatial focusing. This so-called *super-resolution* phenomenon is due to the ability of time reversal to take into account all of the different sound paths between the emitting antenna and the focal point, thus creating an adaptive spatial and temporal matched filter for the considered propagation medium. Experiments emphasize the strong robustness of time-reversal focusing towards small modifications in the medium, such as people in motion or temperature variations. Sound focusing through walls using the time-reversal approach is also experimentally demonstrated. [DOI: 10.1121/1.1543587]

PACS numbers: 43.60.Gk, 43.38.Hz, 43.55.Br, 43.60.Dh [RLW]

I. INTRODUCTION

Most challenging applications in audible range acoustics, such as virtual sound space creation and active noise control, rely on the ability of the rendering system to recreate precisely a desired sound field.[1] This accuracy in rendering has to be achieved both temporally and spatially in a finite volume around the desired listener. From a theoretical point of view, the ability to recreate such a desired sound field is directly linked to the capabilities of the system in terms of spatial and temporal focusing. The propagation medium in practical situations is *a priori* unknown and complex (e.g., multiple reverberations, absorption effects, etc.) thus requiring focusing to be adaptive. Current approaches to achieve this adaptive sound control consider modeling of the propagation medium. Other signal-processing approaches based on the solution of the inverse problem have also been proposed to achieve this focusing objective.[2]

Adaptive focusing of acoustic waves is also an important research topic in domains such as medical imaging and therapy, underwater acoustics, and nondestructive evaluation. The time-reversal method for such applications has been proven a very efficient and easy technique to achieve adaptive focusing regardless of the complexity complex the propagation medium.[3] The aim of this paper is to study both theoretically and experimentally the potential of this technique for audible range acoustics.

The time-reversal focusing technique is based on the time-reversal invariance of the wave equation. In the case of room acoustics, the wave equation can be written in a fairly simple form: Propagation occurs in a lossless fluid with a constant sound velocity c in a medium bounded with nondissipative boundaries (in an ideal case). Thus, for the acoustic pressure field $p(\mathbf{r},t)$ in a transient regime, the propagation of sound obeys the condition[4]

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0. \tag{1}$$

This equation is time-reversal invariant because it contains only second-order derivatives with respect to time. Consequently, for each burst of sound $p(\mathbf{r},t)$ diverging from a source which can then possibly be reflected, refracted, or scattered within the propagation medium, there exists in principle a set of waves $p(\mathbf{r},-t)$ that retraces precisely all of these complex paths and converges simultaneously at the original source site as if time were running backwards.

By taking advantage of the Huygens principle, this time-reversal wave can be obtained by measuring the field $p(\mathbf{r}_S,t)$ on a surface enclosing the experimental volume. This surface is covered with microphones that detect the field $p(\mathbf{r}_S,t)$ during a time T sufficiently long for the wave to vanish. Once this field is stored in memories, the microphones are then replaced by loudspeakers and the surface of loudspeakers reemits the time-reversed signals $p(\mathbf{r}_S,T-t)$. The resulting wave field converges optimally towards the initial source exactly as if the scene was played backwards.

In nondissipative media, such a time-reversal cavity acts as an inverse filter of the diffraction transfer function that relates the wave-field propagation from the source to the closed surface: The movie of the diverging wave propagation is played backwards resulting in an optimal focusing at the initial source location. However, this filter is not perfect because evanescent waves[5] emitted by the source (high spatial frequency content of the wave field) cannot be recorded on

[a]Electronic mail: michael.tanter@escpi.fr

the surface and thus cannot be time reversed. Diffraction acts as a low-pass filter of the spatial frequencies, and it leads to the classical diffraction limit that prevents the size of the time-reversed focal spot from being smaller than half the wavelength. Thus, details of the source smaller than the shortest wavelength are lost during a time-reversal experiment. So, for an ideal point source emitting a wideband pulse, the ''returning'' field refocuses on it with a spot whose dimensions is on the order of the smallest wavelength. For the same reasons, the transducers covering the cavity surface do not need to be uniformly distributed over the whole surface. Because the smallest details are filtered by diffraction, the surface may be sampled by a finite number of transducers distributed on a two-dimensional array with spacing equal to half the smallest wavelength.

This time-reversal cavity remains an idealized concept which is difficult to implement in practice. The strongest limitation is linked to the difficulty of surrounding the focal region by a huge set of transducers. In ultrasound-based medical or nondestructive evaluation applications, a time-reversal mirror (TRM) consisting of a simple linear array is generally used. However, in most cases, its limited aperture reduces the focusing capabilities of the technique.[3,6]

In room acoustics, propagation occurs in a closed cavity with reverberating boundary conditions. This specific configuration leads to interesting properties that can be anticipated from previous results obtained in an ultrasonic frequency range. Indeed, in lossless ultrasonic waveguides,[7] Roux *et al.* have shown that time reversal benefits from the multiple reflections that occur during propagation, permitting the creation of a virtually infinite transducers array. Consequently, the focal resolution of the array is improved in comparison with the focal resolution of the system in free space. For the case of a closed and chaotic 2D cavity,[8] Draeger *et al.* have even shown that the information contained in the reflections of the wave field coming from a single source is sufficient to recreate a complete virtual time-reversal cavity. Thus, a single loudspeaker time-reversal mirror is sufficient to achieve focusing in such a closed and chaotic cavity. Our configuration in room acoustics is slightly different because the studied cavity does not have a chaotic shape and is not as reverberant as the one studied in Ref. 8. However, similarities with the work done by Draeger *et al.*[9] and De Rosny *et al.*[10] can be established. See Fig. 1.

Our goal is to compare the performances of classical time-delay-law focusing and time-reversal focusing in room acoustics. This study is divided into three sections. In Sec. II, time-reversal processing is studied and experimentally achieved using a loudspeaker array inside a reverberating room. The quality of focusing is studied both spatially and temporally and the phenomenon of super resolution achieved by time reversal in the room is clearly demonstrated. In Sec. III, the loudspeaker array is placed outside of the room where focusing is to be achieved. This second experiment allows study of the robustness of the time-reversal focusing process in nonideal conditions where the attenuation due to the walls causes one to reconsider the initial assessment of time-reversal invariance. Finally, the digital communications experiment described in the last section will allow us to quantify the capacity of time reversed to improve the quantity of information that can be transmitted in a reverberating medium.

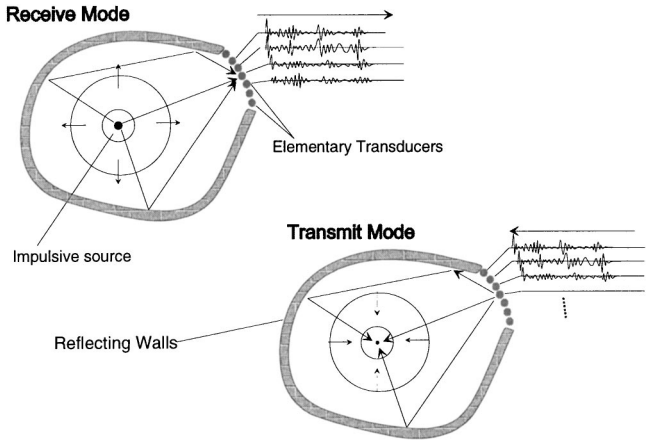


FIG. 1. Time-reversal principle in a reverberating cavity.

II. TIME REVERSAL IN A REVERBERATING ROOM

A. Experimental setup

Experiments have been carried out in a typical room. The room is neither a white reverberating room nor an anechoic room. By choice, we use a room where many people work on a variety of experiments and, consequently, with many obstacles to sound propagation such as tables or computers. This will prove helpful to study robustness of the different tested focusing methods. The typical applications of interest concern speech; for this reason, our study is limited to the bandwidth in the audible frequency range [300–4000 Hz].

The room in which experiments are undertaken is rectangular, with dimensions $4.7\times5.9\times3.1$ m. The focusing system consists of a 20-loudspeaker linear array (Fig. 2), parallel to the measurement bench. The receiving system consists of an electret microphone that scans the pressure field gen-

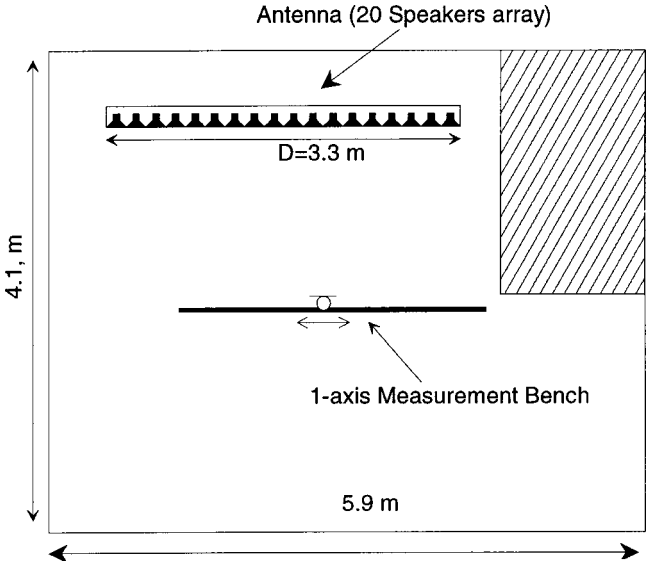


FIG. 2. Setup for experiments in reverberating room.

erated by the antenna along its measurement axis (see Fig. 2). This simple configuration has been deliberately chosen to avoid problems due to modification of the sound field by the measurement system. Emitting and receiving systems are controlled from an electronic device including 12-bit A/D converters for microphones, 256 kBytes memory for each of the 50 time-reversal channels, and 12-bit D/A converters and amplifiers for each loudspeaker. Sampling frequency is set to 20 kHz. This sampling frequency limits us to work in the 0–5-kHz frequency range because focusing quality is linked to the achievable precision in phase between each channel.[11]

Our goal in these experiments is to compare the performances of classical time-delay-law focusing and time-reversal focusing. For practical reasons, there are no active sources at the desired focal point contrary to the usual time-reversal experiment. Indeed, the first step of the time-reversal experiment is conducted using the spatial reciprocity property of the medium: instead of acquiring Green functions $G(\mathbf{Y}_i,\mathbf{r}_{fp},t)$ by emitting from the desired focal point $\mathbf{r}_{fp}$ and receiving at location $\mathbf{Y}_i$ of each transducer of the array, we acquire Green functions by letting each loudspeaker emit individually assuming that $G(\mathbf{Y}_i,\mathbf{r}_{fp},t)=G(\mathbf{r}_{fp},\mathbf{Y}_i,t)$. This method avoids having to use a complex configuration of transducers consisting of a couple of coaxially mounted loudspeaker and microphone for each element of the array.

In the audible range, electro-acoustical response of transducers cannot be neglected, especially for loudspeakers. Impulse responses from an element i of the array to the focal spot are measured by emitting a chirp of bandwidth 100–4100 Hz with the loudspeaker i and acquiring the corresponding signal with a microphone located at the focal point $\mathbf{r}_{fp}$ (the use of a chirp signal excitation instead of a delta function is only due to sensitivity considerations). Impulse responses are then obtained by correlating the signal with the initial chirp, and is given by

$$h_{i,fp}(t)=h^{\mu}_{\mathrm{AE}}(t)^{*}_{t}G(Y_i,r_{fp},t)^{*}_{t}h_{\mathrm{EA}}(t), \tag{2}$$

where $h_{\mathrm{EA}}(t)$ is the electroacoustic impulse response of loudspeaker i (it will be considered identical for each loudspeaker in the present paper), $\mathbf{Y}_i$ is the position of the ith loudspeaker, $h^{\mu}_{\mathrm{AE}}(t)$ is the acousto-electric impulse response for the microphone, and $^{*}_{t}$ denotes convolution with respect to time. Due to the linearity of the problem, impulse response of the electronic system and transducers will be considered as a whole, denoted $h^{i}_{E}(t)$. Thus

$$h_{i,fp}(t)=G(Y_i,r_{fp},t)^{*}_{t}h^{i}_{E}(t). \tag{3}$$

These impulse responses are time reversed and reemitted in the medium by the set of loudspeakers, using the same electronic system. The signal received at location $\mathbf{x}$ in the medium can be written

$$h_{RT}(x,t)=\sum_{i=1,\ldots,N}G(Y_i,x,t)^{*}_{t}G(r_{fp},Y_i,-t)^{*}_{t}\times h^{i}_{E}(t)^{*}_{t}h^{i}_{E}(-t), \tag{4}$$

where $\mathbf{x}$ is a point in the medium. In the case $\mathbf{x}=\mathbf{r}_{fp}$ (i.e., at the focal point) and $t=0$, $h_{RT}(x,t)$ is maximized: Eq. (3) illustrates the fact that time reversal acts as a spatial and temporal matched filter, both for electronic response $h_E(t)$ and for the propagation operator $G(\mathbf{Y}_i,\mathbf{r}_{fp},t)$. This property is emphasized in the next two parts.

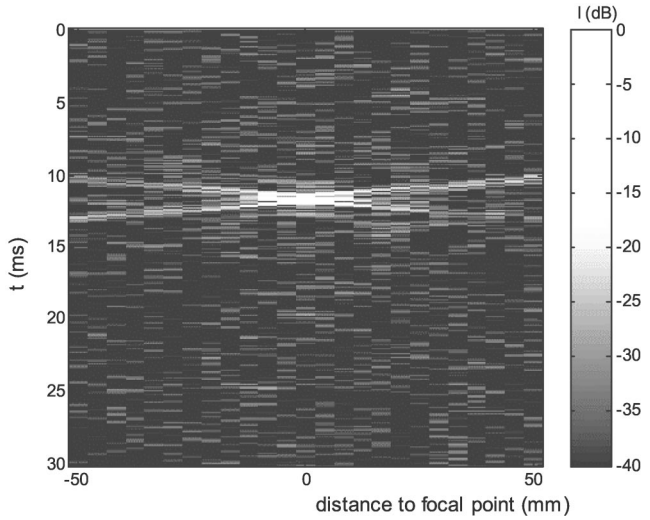


FIG. 3. Intensity of the spatio-temporal focusing pattern achieved by time-reversal processing.

B. Spatial focusing

The performances of the two focusing techniques can be compared using spatial domain point spread functions (PSF), defined as $d(\mathbf{x})=\max_t\{h(\mathbf{x},t)\}$, where $h(\mathbf{x},t)$ is the impulse response obtained at location $\mathbf{x}$ after the focusing process. To obtain the PSFs for the different cases, initially, the set of impulse responses $h_{i,fp}(t)$ relating the loudspeakers noted i and the focal point f_p is acquired. First, the classical focusing technique (time-delay law) is realized by deducing from the experiments the travel time between each loudspeaker and the microphone located at the focal point. Focusing is then achieved by emitting the same signal on each loudspeaker with different delays corresponding to inverse of the previous travel time-delay law. Thus, the signal emitted by each loudspeaker arrives at the same time at the focus [in practice, this time-delay law is computed using the position of the maximum of correlation between $h_{i,fp}(t)$ and $h_E(t)$]. Second, signals to be emitted for time reversal are easily deduced by reversing the set of the $h_{i,fp}(t)$ impulse responses with respect to time.

The set of impulse responses $h(\mathbf{x},t)$ obtained in the focal plane after the focusing process is measured by translating the microphone along the $\mathbf{x}$ axis (see Fig. 2). A typical spatio-temporal response obtained along the focal axis $\mathbf{x}$ by time-reversal processing is presented in Fig. 3.

1. Focal spot width

Figure 4 presents the PSF around the desired focal point obtained with the partial antenna. By comparing PSFs obtained by time reversal with that obtained by the classical delay-law focusing technique, the improvement introduced by time reversal is evident: the main lobe is sharper, and sidelobes are 7 to 10 dB lower.

The dashed intermediate curve corresponds to the simulation of time-reversal focusing when propagation occurs in

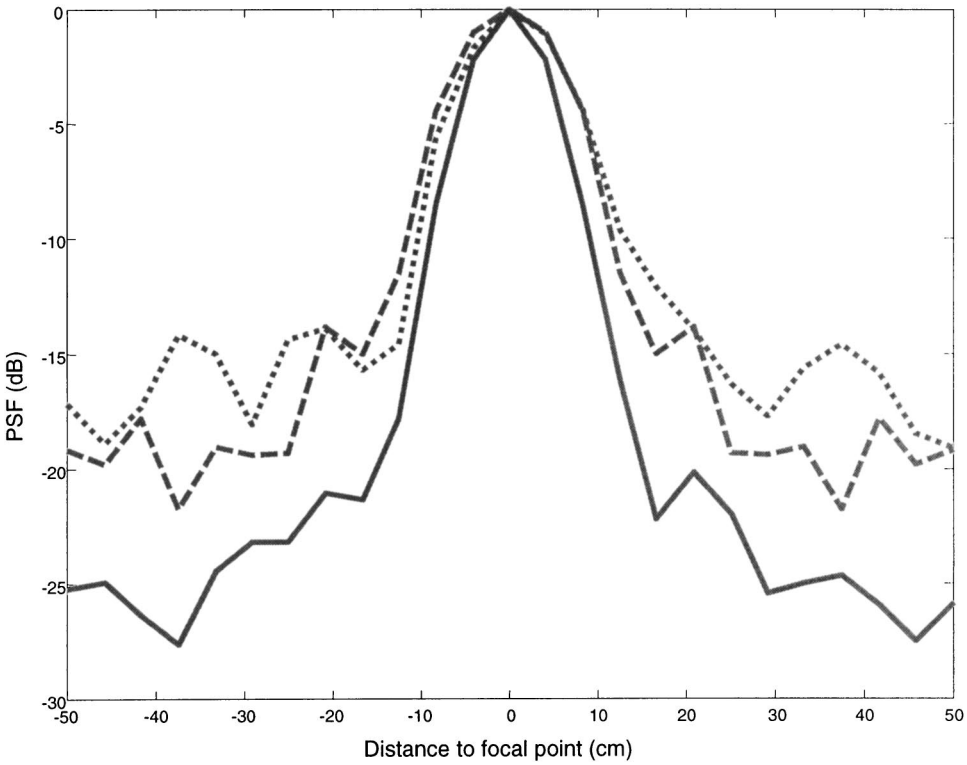


FIG. 4. Point spread functions obtained with the partial antenna: Delay law (dotted line), TR in free space simulation (dashed line), and experimental TR in the room (solid).

free space. This result is obtained by assuming that transducers are point-like. In this situation, Green's function relating to point **x** and **y** can be written as

$$G(\mathbf{x},\mathbf{y},t)=\frac{1}{\|\mathbf{xy}\|}\,\delta\left(t-\frac{\|\mathbf{xy}\|}{c}\right),$$

where $\|\mathbf{xy}\|$ is the distance between points **x** and **y**, and $\delta(t)$ is the delta function. Substitution of this formula in Eq. (3) leads to

$$h_{RT}(x,t)=\sum_{i=1}^{N}\frac{1}{\|xY_i\|\|r_{fp}Y_i\|}\times\delta\left(t-\frac{\|xY_i\|-\|r_{fp}Y_i\|}{c}\right)^{*}_{t}h_E^i(-t)^{*}_{t}h_E^i(t). \tag{5}$$

A good approximation of the impulse response of electronic system and transducers $h_E(t)$ is measured in an anechoic room using the actual system; calculation of $h_{RT}(\mathbf{x},t)$ is then straightforward. The corresponding dashed curve in Fig. 4 demonstrates the advantages of time-reversal process. Time reversal in free space provides results that are very similar to those obtained with the classical delay-law focusing. The slight improvement provided by time reversal in such a situation is due to the temporal matched filtering of the impulse response of the electronic system $h_E(t)$. In the case of propagation in a reverberating medium, such as our room, time reversal gives results that appear to be even better than those obtained in free space.

This phenomenon, called *super-resolution*, is also observed in waveguides[7] and in closed cavities.[8] After time reversal, the rays that arrive at the focal point appear to be emitted from multiple sets of virtual sources. These sources are the images of the actual emitting antenna with respect to the different interfaces in the rooms (walls or furniture). The superposition of the signals coming from each virtual antenna at the focal point leads to a virtually infinite antenna. Thus, focusing by time reversal in a highly reverberant room can be assimilated to near-field focusing, and the focal spot width does not depend on the size of the antenna, but is always equal to $\lambda_0/2$, where λ_0 is the central wavelength of the used signal. In the present experimental case, we can clearly check that the focal spot has a lateral width of $\lambda_0/2 = 17$ cm, corresponding to the central frequency of the loudspeakers, 975 Hz.

In Fig. 5 the size of the focal spot is plotted versus the

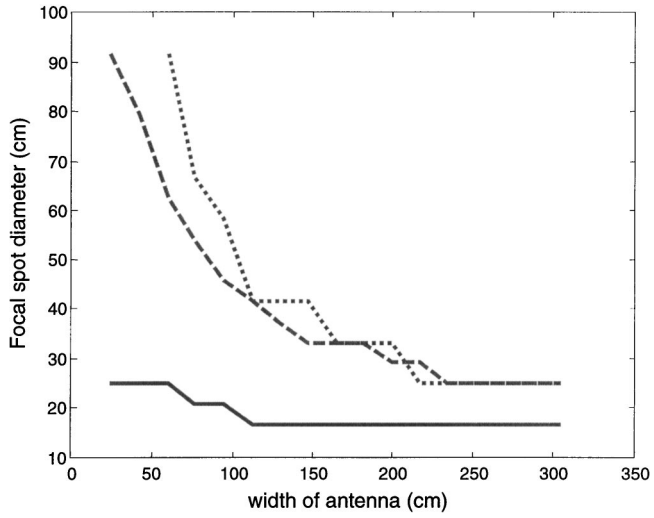


FIG. 5. Focal spot width vs antenna diameter for the delay law (dotted line), TR in free space simulation (dashed line), and TR in the room (solid).

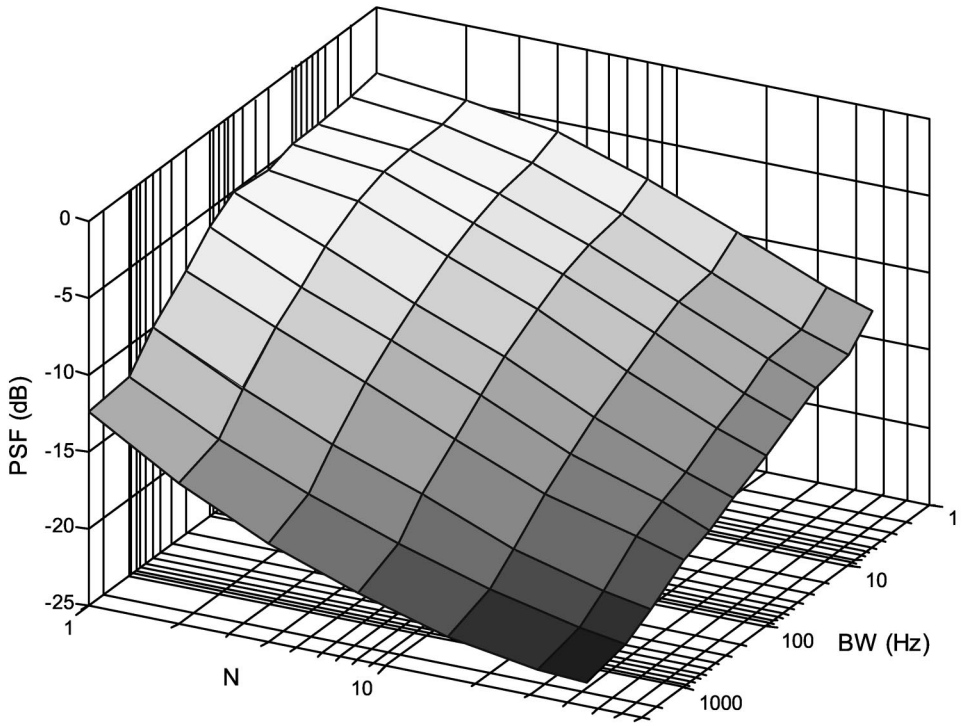


FIG. 6. Sidelobe level $SL_{(dB)}$ as a function of the number of used loudspeakers N and of the bandwidth BW. Value is the average of 32 repetitions.

size of the antenna for the different focusing techniques. As stated above, the size of the focal spot remains constant for time-reversal focusing in a reverberating room corresponding to the optimal focus size $\lambda_0/2$. In the case of focusing in free space, the focal spot width follows the classical diffraction relation $d=\lambda_0(F/D)$, where F is the distance between antenna and focal point, and D the diameter of the antenna. One can clearly notice that the super-resolution phenomenon is experimentally checked as soon as the antenna is larger than $3\lambda_0$ (i.e., 100 cm). The slight variations observed for smaller antenna diameters on the focal spot width obtained by time-reversal processing show the limits of the virtual antenna approach: the virtual antenna is infinite if the loudspeakers are omnidirectional and if no attenuation occurs during reflections, which is not the case in reality. After each reflection on the wall, the amplitude of the reverberated signal is smaller and it results in a corresponding virtual source of lower efficiency.

2. Secondary lobes level

Beside the importance of obtaining a narrow focal spot, sidelobe level is the second parameter to take into account when evaluating the performance of a time-reversal focusing system. This parameter has been studied by Roux *et al.*[7] for the case of a waveguide, where a simple interpretation of the use of reflections by time reversal in terms of virtual images is introduced. In this article, sidelobe levels are shown to be dependent on the number of transmitters and on the dimensions of the waveguide. More complete analysis is available in Ref. 5. In the present case, other parameters require consideration, such as the available bandwidth or the size of the temporal window used for time-reversal processing. The subjective effect of sound focusing for the human ear is mainly linked to two parameters: quality of sound at the focal point (which depends on the quality of the impulse response after focusing), and the relative level of the first sidelobe (dynamic of the focusing). Thus, we introduce the sidelobe level as $SL_{(dB)}=\langle 20\log_{10}[d(y)/d(0)]\rangle$, where $d(y)$ is the PSF and y varies between $\lambda_0/2$ and $2\lambda_0$.

Figure 6 represents the evolution of $SL_{(dB)}$ with the number of loudspeakers and the bandwidth used during the time-reversal experiment. In order to understand this evolution, we consider four aspects of the function SL (N,BW).

(i) In the center of the figure, $SL_{(dB)}$ evolves in a bilinear way: $SL_{(dB)}=A\log_{10}(N)+B\log_{10}(\mathrm{BW})+C$, where N is the number of loudspeakers, BW the effective used bandwidth, and A, B, C are constants. This point can be explained considering that, due to their different reverberations, the contributing signals coming from each loudspeaker are uncorrelated; thus, they will sum coherently at the focal time, and incoherently elsewhere, leading to a ratio: $d(y)/d(0)\propto\sqrt{N}$. In frequency domain, the number of excited modes is proportional to the bandwidth. Similarly, it may be considered that modes are uncorrelated and contribute incoherently except at focal point. Here again, this leads to a relation: $d(y)/d(0)\propto\sqrt{\mathrm{BW}}$. This bandwidth dependence was addressed by Draeger *et al.* for the case of chaotic cavities.[8,9] Substitution in the above formula gives: $St_{(dB)}=10(\log_{10}(N)+\log_{10}(\mathrm{BW}))+C$. This simple approach provides a coarse approximation of the surface. In fact, relation between bandwidth and number of excited modes appears to be more complicated.

(ii) When bandwidth becomes very small, the excitation is quasimonochromatic. In this case, point spread function in free space is of the form $d(y)=[\sin(2\pi\lambda/Dy)]/[2\pi\lambda/Dy]$. This PSF leads to a

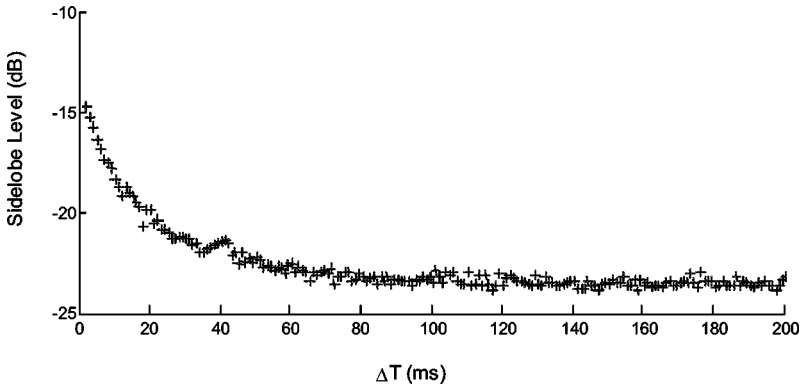


FIG. 7. Dependence of sidelobes level $SL_{(dB)}$ on quantity of data ΔT used for time reversal.

minimum sidelobes level of -13 dB; this agrees with the minimum value of $SL_{(dB)}$ obtained experimentally when N increases for small bandwidths.

(iii) For large bandwidths, $SL_{(dB)}$ seems to reach a limit. This an artifact of the experiment, corresponding to the fact that one bandwidth is limited by the electronic system itself, and not by the applied filter.

(iv) When $N=1$, it can be noticed that $SL_{(dB)}$ reaches a maximum of -4 dB as the bandwidth decreases. Thus, even a single loudspeaker is able to achieve focusing by using the multiple reverberations in the room. This situation is explained with a simple model: the virtual images of the loudspeaker with respect to the different interfaces in the room allows creation of a spatially undersampled antenna. The focusing pattern corresponds to the focusing quality obtained with such an undersampled antenna with monochromatic excitation in free space. A simple model is established considering that time reversal in this situation creates a virtual antenna of transducers with a step equal to the width of the room (5.9 m). The PSF is then calculated using the Huygens–Fresnel principle.[11] In conclusion, even with one loudspeaker and a monochromatic excitation, a minimal focusing does exist as the time-reversal approach benefits from the multiple reflections.

Another interesting point is the quantity of data needed to achieve a good quality of focusing. In Fig. 7, $SL_{(dB)}$ is represented against the size of the temporal window used during the time-reversal process. This figure denotes the evolution from small ΔT (where focusing is very similar to the one obtained with a delay-law technique) to large ΔT for which all available reflections are taken into account in the time-reversal operation. Draeger *et al.*[8] show the existence of a saturation level for sidelobes in regards to ΔT, linked to the fact that not all the modes of the cavity can be equally excited and that this saturation time can be related to the Heisenberg time. However, the saturation time observed in our case cannot be interpreted this way: First, our cavity is

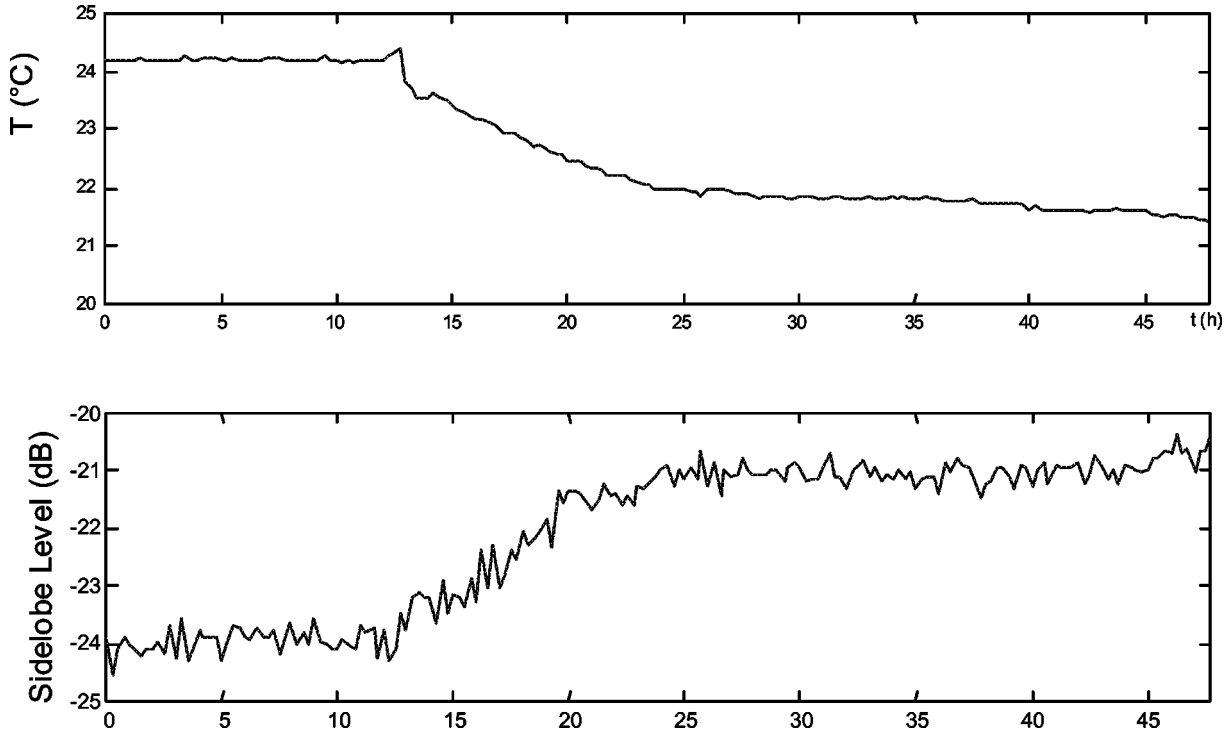


FIG. 8. Evolution of sidelobe level SL(dB) over a 48-h period (bottom) and the corresponding temperature evolution (top).

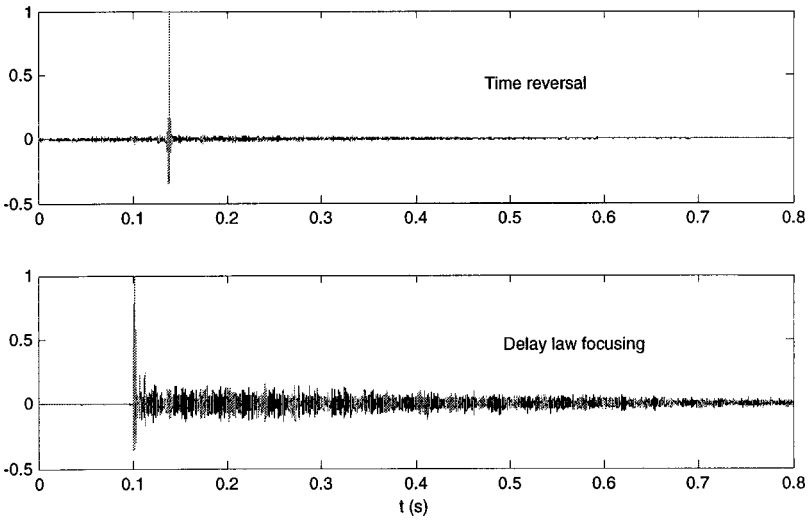


FIG. 9. Temporal signal received at focus for delay-law focusing (bottom) and for time reversal at the focal point $h_{RT}(r_{\mathrm{fp}},t)$.

not strictly speaking chaotic. Moreover, observation of such a saturation phenomenon requires a very low attenuation during reflections and signal-to-noise ratios that are impossible to obtain in a standard room and generally in audible acoustics contrary to the silicon wafers used in Ref. 8. In our case, emitted signals are shorter due to attenuation: thus, the observed saturation is due mainly to the signal falling below the noise level.

3. Influence of changes in the propagation medium

In practical systems, focusing needs to be robust: focusing has to be achieved in media where objects will eventually move, and conditions of the propagation can vary. The influence of such an evolution upon focusing is measured by acquiring an initial set of signals and calculating the corresponding set of time-reversed signals to emit. This set is then reemitted every 15 min in order to study the evolution of the sidelobe level $\mathrm{SL}_{(\mathrm{dB})}$ against time. Results are presented in Fig. 8. During the measured period (48 h), people move and work in the room, creating slight modifications of the propagating medium. In these conditions where temperature remains constant (variations do not exceed 0.5 °C), focusing through time reversal proves to be very robust. However, time-reversal focusing is very sensitive to cumulative variations of the propagation medium characteristics. Figure 8 shows that for example focusing quality has a strong dependence with temperature (i.e., with variations of sound speed). This problem can easily be overcome in practical applications by using banks of data corresponding to the different conditions and link the choice of the data bank to the temperature in the room.

C. Temporal focusing

In the time domain, results presented in Fig. 9 show the effect on time reversal for the compression of the impulse response of the room. As shown in Eq. (2), the response obtained at the focal point corresponds to the coherent sum of the autocorrelation of the impulse responses for each loudspeaker. For this reason, the response is symmetric and is a good approximation of a delta function. However, the shape of the temporal response at the focal point for the two focusing techniques appears to be very different: Time reversal optimizes the ratio between the main focusing lobe and sidelobes in a symmetric way, whereas delay-law techniques concentrate noise after the main lobe, giving a noisier but more natural response, since it is similar to what would be obtained for the propagation of a single pulse in the room.

Thus, from the temporal point of view, time reversal appears to give an interesting result in term of S/N ratio. This result is directly linked to the fact that time reversal provides both a temporal[12] and spatial[13] matched filtering, corresponding to the real experimental conditions. However, the symmetrical form of the impulse response at the focal point, although optimal in terms of S/N ratio, makes the system difficult to use for audio applications. Indeed, the sidelobes preceding the main lobes create strange perspective effects,

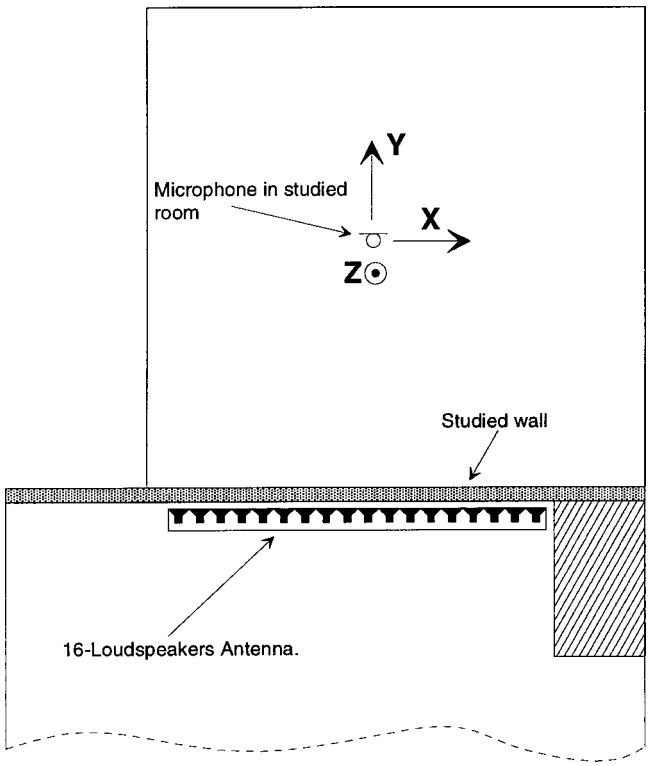


FIG. 10. Experimental setup for focusing through the wall.

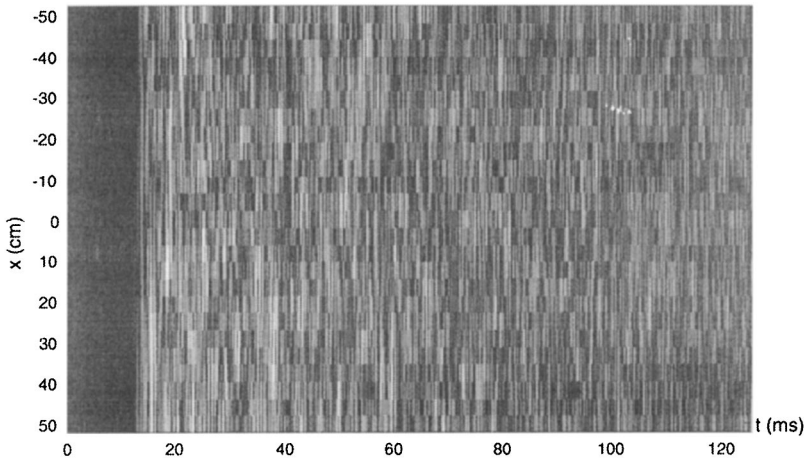


FIG. 11. B scan for the signal coming from loudspeaker No. 8: $h_8(x,t)$.

especially with percussive sounds and plosives consonants, whose rendering quality is poor.

Experiments conducted in the room demonstrated the efficiency and robustness of time reversal compared to the classical time-delay technique to achieve focusing in a reverberant room, both in terms of temporal compression and spatial focusing. A more challenging configuration for time reversal is investigated in the next section.

III. TIME REVERSAL THROUGH A WALL

Focusing sound in a reverberant room appears to be a nearly ideal situation for time reversal: Attenuation is due mainly to reflections, and remains quite small. In order to study the robustness of the time-reversal process for audible range acoustics, the focusing experiment described in the preceding section is now conducted under the conditions where attenuation is strong and propagation much more complex. The experimental setup is described in Fig. 10: The 20-loudspeaker antenna array is now directed towards the wall and we try to focus through the wall in the other room.

Propagation of acoustical energy from one room to the other is complex, because sound propagates through multiple paths, such through the wall, of course, but also through doors and windows. As shown in Fig. 11, there is no direct wavefront, and no specific path for propagation can be identified. In the conditions of experiment, the acoustic attenuation (acoustic attenuation is defined as the difference in mean sound levels in each room for white noise) between the central loudspeaker of the antenna and the central position of the microphone is 20 ± 3 dB. The wall by itself is made of hollow bricks and plaster, and is 8 cm thick. Such a wall provides a normalized acoustic attenuation of $\sim$41 dB.[14] Thus, the most important part of acoustical energy that manages to reach the focal point comes from ''leaks,'' probably via doors and windows but also from lateral propagation through walls. The problem is nevertheless very interesting because the attenuation is strong and strongly frequency dependent, as shown in Fig. 12. (For example, some cuts are down 30 dB.)

Those conditions are very tricky for time reversal. Indeed, its ability to focus should be degraded by attenuation, as the technique is based on the fact that no dissipation occurs during propagation. Moreover, time reversal does not provide any compensation for the amplitude of the frequency response, and this point will necessarily lead to reductions in the focusing quality.

A. Focusing experiments

A typical PSF obtained in the configuration described above is plotted in Fig. 13. The results obtained in such a configuration are very similar to those obtained in a simple room, especially in terms of focal spot width: -6-dB width is equal to $\lambda/2$, where λ corresponds to the central frequency of the loudspeakers. On the other hand, secondary lobe level is higher (around -18 dB) than the one obtained for an equal number of loudspeakers in the room. This phenomenon is

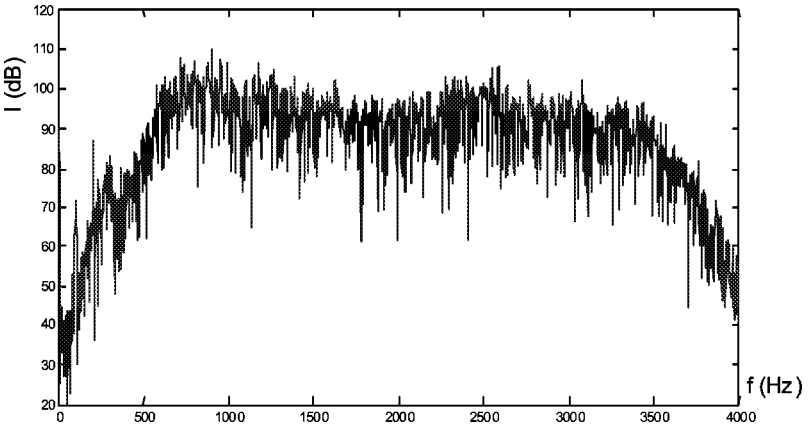


FIG. 12. Spectral density of the impulse response $h_8(r_{fp},t)$ between loudspeaker No. 8 and the desired focal point.

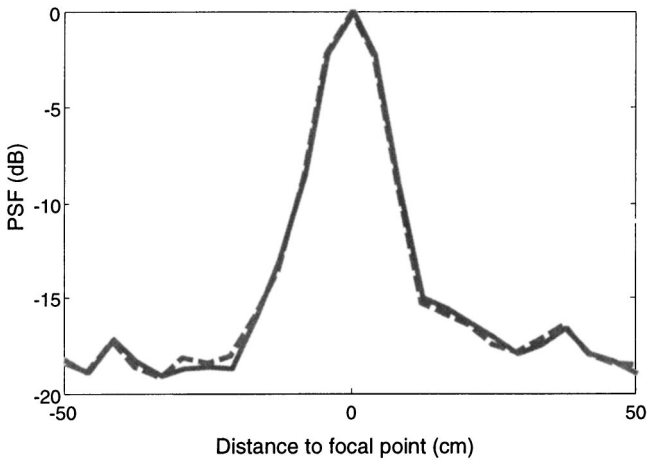


FIG. 13. Point spread functions obtained with time reversal (dashed line) and time reversal with amplitude compensation (solid line).

linked to acoustical dissipation during propagation, which implies distortion in the spectrum and distortion of the overall amplitude coming from each loudspeaker. Figure 14 shows the consequences of these distortions for the overall quality of focusing: In many cases, no focusing can be achieved (sidelobe level is 0 dB or higher), whereas in similar conditions, focusing was possible when using the first experimental setup (see Fig. 6). Thus, loss of information during propagation results in less quality in spatial focusing, because the constructive interference system at the focal point is less effective.

From the point of view of temporal compression, results are very similar to those obtained with time reversal with the conditions of the first experimental setup, except that the temporal sidelobe level is higher: The level of first sidelobes cannot be lowered under -25 dB, whereas levels of -35 dB could be attained in rooms. In the meantime, as for spatial behavior, the width of temporal response is equal to the width obtained in rooms.

Previously, we have studied focusing only along one axis. But, the shape of the focal spot can be studied in three dimensions. Since time reversal benefits from the multiple

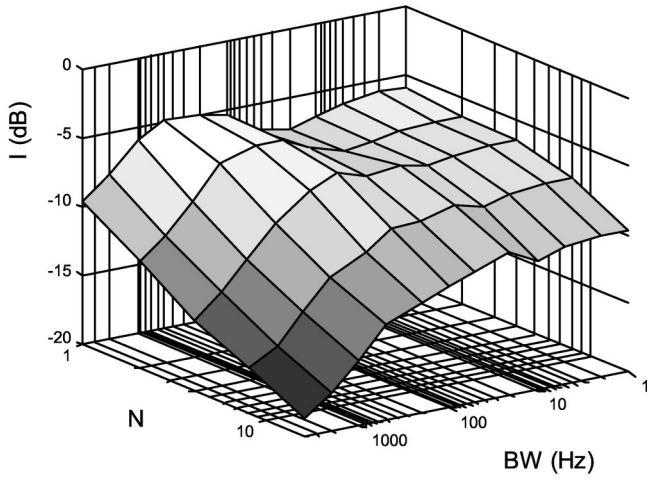


FIG. 14. Sidelobe level $SL_{(dB)}$ as a function of the number of used loudspeakers N and of the bandwidth BW. Value is the average of 32 repetitions.

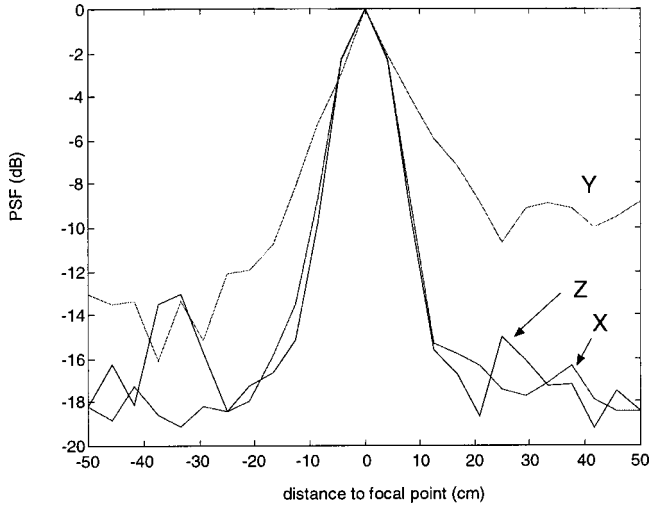


FIG. 15. Point spread functions for focusing through the wall along the three axes $\{x,y,z\}$.

reflections in the three dimensions, we hope to obtain a similar focusing whatever the measurement axis is. Moreover, focal spot width should remain the same, equal to $\lambda/2$, where λ corresponds to the resonance frequency of the loudspeakers. Sidelobe levels can vary because the virtual images of loudspeakers created by reflections leads to spatially undersampled virtual antennas.

Figure 15 presents the PSFs obtained along the three dimensions $\{x,y,z\}$: As suggested above, focusing occurs in the three dimensions. One should notice that vertical and lateral focusing (axes **x** and **z**) are similar and optical ($\lambda/2$ spot size), although the antenna is parallel to the **x** axis. However, longitudinal focusing proves to be more difficult to achieve, probably because of attenuation during reflections: In this direction, focusing occurs indirectly through reflections against the opposite wall in the other room, so more reflections are needed and attenuation is stronger.

We note that similar results are obtained when the active sources are placed directly in the room. These are not presented to avoid redundancy, and because the present situation is more challenging for such an experiment.

B. Communication experiment

The quality of focusing obtained with time reversal through walls is too poor to be able to transmit music with good quality from one room to the other, especially because of important variations in the frequency response of the system, and because of the high level of the temporal sidelobes. However, this technique can prove very useful in a context of digital communications. For example, using the previously described experimental setup, we can transmit information in a binary form.

To achieve this transmission, a classical method of modulation and demodulation is used. In order to concentrate the emitted signals in the frequency range of the focusing system, information is modulated around a carrier frequency of 2500 Hz, using a Q-DPSK technique.[15] As it is shown in Fig. 16, measurement imprecision can be quite important both for phase and amplitude, justifying the choice of such a

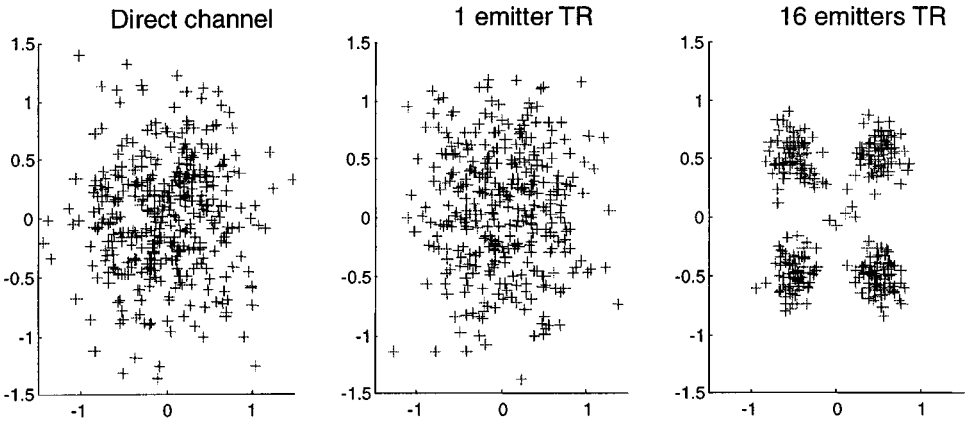


FIG. 16. Signal-space representations of the symbols acquired at the desired receiving location.

modulation system. Furthermore, the fact that modulation is differential provides an easy way to get rid of problems with baud clock synchronization.

In order to evaluate the performances of time reversal for information transmission under such circumstances, three different filtering techniques are used: First, the signal $s(t)$ obtained from the modulator is transmitted as is with one of the loudspeakers. A second technique consists of emitting the $s(t)$ signal with one loudspeaker after filtering with the time-reversed propagation response between the loudspeaker and the microphone. Thus, we obtain a temporal matched filtering of the propagation medium with a single loudspeaker. Finally, $s(t)$ is transmitted after filtering by the different time-reversal responses between each loudspeaker and the microphone. Figure 16 shows how these different transmission techniques act on the signals constellations. When compared to direct transmission or single channel matched filtering, time reversal for multiple channels provides an improvement in both phase and amplitude measurements due to the fact that the improvement given by matched filtering is coherently summed for each of the propagation paths.

Quantitative results about this experiment are obtained by simulating the transmission a succession of random bit sequences, each of them containing 25 kbits, at a band rate of 1250 symbols/s. A bit error rate (BER) is then calculated by comparing the obtained sequence and the transmitted sequence. Each measurement is repeated 512 times and the BERs are finally averaged to obtain the results of Table I. In the case of direct transmission, no detection is possible (BER$\cong$50%): This point is directly linked to the fact that no equalization of the channel is provided.

From these results, time reversal proves to be an efficient way to provide channel equalization and to use channel diversity. Indeed, Table I shows that a simple matched filter (case of a one-loudspeaker antenna) is not sufficient in our case of interest, whereas the use of a multiple-elements antenna allows transmission to occur under varied conditions. Thus, this technique can be very useful in environments where propagation creates multiple reflections and complex distortion of the channel, even when strong attenuation is present.

TABLE I. Evolution of the bit error rate (BER) with the number of loudspeakers in the transmit antenna.

Number of elements in antenna	1	2	4	8	16
BER	47.3%	12.7%	0.12%	3.10^{-4}	$<10^{-5}$

IV. CONCLUSION

We have presented in this paper experimental results on time-reversal properties in audible range acoustics. Thanks to an optimal use of the reverberated field, due to its spatial and temporal matched filter property, time reversal achieves high focusing quality, both from the temporal and spatial points of view. In particular, by using the time-reversal approach, a loudspeaker antenna benefits from the reverberations in the room to achieve a better focal spot than in free space. This phenomenon, known as super-resolution, is experimentally demonstrated here for room acoustics. So, according to its spatial and temporal focusing ability, time reversal could be very useful to control the local sound field in a 3D volume around a focal point, even if the focal point is not in the same room as the control antenna. This technique may find numerous applications in different transmission systems, such as virtual source imaging systems for multimedia applications or devices for digital communications in complex media, as shown in the previous section.

However, its potential remains limited for audio applications due to temporal sidelobes preceding the main signal that create strange perspective effects, especially with percussive sounds and plosives consonants, whose rendering quality is quite poor. Moreover, time reversal proves to be sensitive to attenuation and spectral content of the system. Thus, as soon as the propagation medium becomes strongly dissipative or if the loudspeaker bandwidth is nonuniform, focusing quality can be degraded. In that case, more sophisticated techniques, such as inverse filtering, as proposed by Tanter[16] or Kahana,[17] can be envisioned and are currently under investigation. The application of inverse filtering to sound focusing is described and compared to the time-reversal approach in following works.

ACKNOWLEDGMENTS

The authors would like to thank S. Conti for his useful contribution to the experiments, and V. Gibiat and K. Waters

for their wise comments on the manuscript. This work was supported in part by a DGA grant (ERS No006001A/A000/DSP/SREA/SC/SR).

[1] A. J. Berkhout, D. de Vries, and P. Vogel, ''Acoustic control by wave field synthesis,'' J. Acoust. Soc. Am. **93**, 2764–2778 (1993).
[2] O. Kirkeby, P. A. Nelson, and H. Hamada, ''Local sound field reproduction using two closely spaced loudspeakers,'' J. Acoust. Soc. Am. **104**, 1973–1981 (1998).
[3] M. Fink, ''Time reversal of ultrasonic fields I. Basic principles,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 555–566 (1992).
[4] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[5] J. D. Maynard, E. G. Williams, and Y. Lee, ''Near-field acoustic holography. I. Theory of generalized holography and the development of NAH,'' J. Acoust. Soc. Am. **78**, 1395–1413 (1985).
[6] M. Fink, ''Time reversed acoustics,'' Phys. Today **20**, 34–40 (1997).
[7] P. Roux and M. Fink, ''Time reversal in a waveguide: Study of the temporal and spatial focusing,'' J. Acoust. Soc. Am. **107**, 2418–2429 (2000).
[8] C. Draeger and M. Fink, ''One channel time-reversal of elastic waves in a chaotic 2D-silicon cavity,'' Phys. Rev. Lett. **79**(3), 407–410 (1997).
[9] C. Draeger, J.-C. Aime, and M. Fink, ''One-channel time reversal in chaotic cavities: Experimental results,'' J. Acoust. Soc. Am. **105**, 618–625 (1999).
[10] J. de Rosny, A. Tourin, and M. Fink, ''Coherent backscattering of an elastic wave in a chaotic cavity,'' Phys. Rev. Lett. **84**(8), 1693–1695 (2000).
[11] J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed. (McGraw-Hill, New York, 1996).
[12] C. Dorme and M. Fink, ''Focusing in transmit–receive mode through inhomogeneous media: The time reversal matched filter approach,'' J. Acoust. Soc. Am. **98**, 1155–1162 (1995).
[13] M. Tanter, J.-L. Thomas, and M. Fink, ''Time reversal and the inverse filter,'' J. Acoust. Soc. Am. **108**, 223–234 (2000).
[14] F. Fahy, *Foundations of Engineering Acoustics* (Academic, New York, 2001).
[15] John G. Proakis, *Digital Communications*, 4th ed. (McGraw-Hill, New York, 2000).
[16] M. Tanter, J.-F. Aubry, J. Gerber, J.-L. Thomas, and M. Fink, ''Optimal focusing by spatio-temporal inverse filter. I. Basic principles,'' J. Acoust. Soc. Am. **110**, 37–47 (2001).
[17] Y. Kahana, P. A. Nelson, O. Kirkeby, and H. Hamada, ''A multiple microphone recording technique for the generation of virtual acoustic images,'' J. Acoust. Soc. Am. **105**, 1503–1516 (1999).

A second, low-frequency mode of vibration in the intact mammalian cochlea

Andrei N. Lukashkin[a] and Ian J. Russell
School of Biological Sciences, University of Sussex, Falmer, Brighton BN1 9QG, United Kingdom

(Received 8 August 2002; revised 7 November 2002; accepted 9 November 2002)

The mammalian cochlea is a structure comprising a number of components connected by elastic elements. A mechanical system of this kind is expected to have multiple normal modes of oscillation and associated resonances. The guinea pig cochlear mechanics was probed using distortion components generated in the cochlea close to the place of overlap between two tones presented simultaneously. Otoacoustic emissions at frequencies of the distortion components were recorded in the ear canal. The phase behavior of the emissions reveals the presence of a nonlinear resonance at a frequency about a half octave below that of the high-frequency primary tone. The location of the resonance is level dependent and the resonance shifts to lower frequencies with increasing stimulus intensity. This resonance is thought to be associated with the tectorial membrane. The resonance tends to minimize input to the cochlear receptor cells at frequencies below the high-frequency primary and increases the dynamic load to the stereocilia of the receptor cells at the primary frequency when the tectorial membrane and reticular lamina move in counterphase. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1535941]

PACS numbers: 43.64.Jb, 43.64.Kc, 43.64.Ri [BLM]

I. INTRODUCTION

The mammalian cochlea functions with precise time resolution as a sensitive, sharply tuned frequency spectrum analyzer. Each point along the cochlear partition is most sensitive to a particular frequency, which is called the characteristic frequency (CF) of the place. These features of the cochlea depend on amplification provided by high-frequency electromechanical feedback (Dallos and Evans, 1995; Frank *et al.*, 1999) from the outer hair cells (OHCs). The OHCs are stimulated through shear displacement (Davis, 1958) between the organ of Corti (OC), which supports the OHC soma, and the extracellular matrix of the tectorial membrane (TM), in which the tips of the OHC stereocilia bundles are inserted. This relative shear displacement might be expected to demonstrate a complex dependence on frequency and level of stimulation because the mammalian cochlea is a composite structure comprising many components that are connected by elastic elements. Consequently, the cochlea is a dynamic system with several mechanical degrees of freedom. The response of each place along the length of the cochlea to incoming sound is likely to be a combination of the normal vibrational modes of the system. Thus, each tonotopic place in the mammalian cochlea might be expected to respond most sensitively at several different frequencies that correspond to resonances at frequencies of the normal modes of vibration.

From recent measurements of the basilar-membrane (BM) mechanical responses made from normal, wild-type, mice and genetically modified mice with detached TM (Legan *et al.*, 2000), there is strong experimental evidence that each section of the cochlear partition resonates at two or more frequencies. In normal mice, the BM iso-response-tuning curve reveals a sharp threshold minimum and associated level-dependent phase changes at the CF. Namely, the phase angle lags or leads with increasing level of stimulation for frequencies below or above the CF, respectively. Accordingly, the phase angle is almost independent of level for frequencies at or very close to the CF of the recording place. These phase changes provide a signature of the nonlinear resonance at the CF place. A second, more broadly tuned threshold minimum and related level-dependent phase changes was recorded about a half octave below the CF. The iso-response-tuning curves of the BM displacements recorded from similar locations in the cochleae of mutant mice without TM have a sharp but less sensitive minimum at the expected CF of the measurement site, but the second low-frequency minimum and accompanying phase changes are absent. This is a strong indication that the TM is a second mass that, together with the rest of the OC, forms a resonating system with at least two degrees of freedom. If this were the case, then the acting stimulus for the hair cells, which is the relative displacement of the TM with respect to the OC, would depend strongly on the mode of vibration of the cochlear partition for specific signal parameters. Therefore, it is of crucial importance for our understanding of the signal processing in the cochlea to identify the modes of vibration for each cross section of the cochlear partition.

With the exception of BM measurements in the mustached bat cochlea (Russell and Kössl, 1999), measurements of BM responses in the hook region of the mouse cochlea (Legan *et al.*, 2000) differ from those made in other species, (see Robles and Ruggero, 2001, for review). To date, a low-frequency resonance at the CF place has been reported only in the mouse and the bat cochlea. This might be because only in these two species have recordings of BM vibrations been made through the transparent round window membrane of the intact cochlea. In other studies of the BM mechanical

[a]Electronic mail: a.lukashkin@sussex.ac.uk

 0001-4966/2003/113(3)/1544/7/$19.00

tuning, the BM was viewed through an opening in the bony wall of the cochlea. Such cochlear surgery has the associated risk of changing the overall mechanical properties of the cochlea as a consequence of modifications to the cochlea's hydrodynamics. Disruption of the cochlear integrity in the apical turn of the cochlea has been reported (Cooper and Rhode, 1996) to lead to significant alteration in its mechanical responses. It is therefore worth attempting to use a noninvasive technique to probe frequency tuning of the *intact* cochlea in a species other than bat and the mouse to see if more than one resonance contributes to the pattern of vibration at a single location on the cochlea partition. In the current study, the cochlear response at the CF place to frequencies well below the CF was examined using low-frequency distortion components.

Mechanical responses of the mammalian cochlea, at least in its basal turn, have a significant level of distortion (Robles *et al.*, 1991; Cooper, 1998). Part of the energy at frequencies of the distortion components is reflected back from the cochlea and can be recorded in the ear canal in the form of distortion product otoacoustic emissions (DPOAE) (Kemp, 1979). DPOAE generation is closely associated with the OHC motility (Frolenkov *et al.*, 1998) at all levels of the primary tones (Lukashkin *et al.*, 2002). Over a wide range of stimulus parameters, DPOAEs originate mainly from the place on the BM where the excitation due to the primary tones overlap, i.e., from the cochlear region, which is close to the f_2 CF place (Brown and Kemp, 1984; Knight and Kemp, 2000). Therefore, if f_2 is kept constant and f_1 changes, then the frequencies of the low-frequency components of the DPOAE will sweep through the frequency region of the hypothetical low-frequency resonance at the f_2 CF place (Russell and Kössl, 1999; Legan *et al.*, 2000). Hence, one can expect that the DPOAE components will demonstrate a level-dependent phase change specific to the nonlinear resonance. Namely, a phase lag of the DPOAE components with frequencies below the postulated resonance should change to phase lead when frequencies of the components exceed the resonance frequency. This transition of the phase behavior should occur at around the same frequency for all frequency components, e.g., for $2f_1-f_2$, $3f_1-2f_2$, $4f_1-3f_2$, etc., in other words for different f_2/f_1 ratios. However, it should be borne in mind that the DPOAE amplitude does not necessarily peak at the TM resonance frequency as it has been suggested earlier (Brown *et al.*, 1992; Allen and Fahey, 1993). According to the latest experimental and theoretical results, the bandpass structure of the DPOAE could have a complex origin (Kanis and de Boer, 1997; Stover *et al.*, 1999; van Dijk and Manley, 2001; Lukashkin and Russell, 2001; Vetešník and Nobili, 2002) and processes other than filtration might contribute to the generation of the bell-like filter shape of the DPOAE. It is also worth noting that the resonance frequency in nonlinear systems depends on the level of the input signals. Accordingly, the frequency of the phase transition could be different for primary pairs of different levels.

II. METHODS

Eleven pigmented guinea pigs (250–350 g) were anesthetized with the neurolept anesthetic technique (0.06-mg/kg body weight atropine sulphate s.c., 30-mg/kg pentobarbitone i.p., 500 μl/kg Hypnorm i.m.). Additional injections of Hypnorm were given every 40 min. Additional doses of pentobarbitone were administered as needed to maintain a nonreflexive state. The heart rate was monitored with a pair of skin electrodes placed on both sides of the thorax. The animals were tracheotomized and artificially respired, and their core temperature was maintained at 38 °C with a heating blanket and head holder. The middle-ear cavity of the ear used for the DPOAE measurements was open to equilibrate air pressure on the both sides of the tympanic membrane (Zheng *et al.*, 1997).

Sound was delivered to the tympanic membrane by a closed acoustic system comprising two Bruel & Kjaer 4134 $\frac{1}{2}$ -in. microphones for delivering tones and a single Bruel & Kjaer 4133 $\frac{1}{2}$-in. microphone for monitoring sound pressure at the tympanum. The microphones were coupled to the ear canal via 1-cm long, 4-mm diameter tubes to a conical speculum, the 1-mm-diameter opening of which was placed about 1 mm from the tympanum. The closed sound system was calibrated *in situ* for frequencies between 1 and 50 kHz. Known sound-pressure levels were expressed in dB SPL $re{:}2\times10^{-5}$ Pa. White noise for acoustical calibration and tone sequences for auditory stimulation were synthesised by a Data Translation 3010 board at 200 kHz and delivered to the microphones through low-pass filters (100-kHz cutoff frequency). Signal from the measuring amplifier was digitized at 200 kHz using the same board and averaged in the time domain. Amplitudes and phase angles of the spectral peaks were obtained by performing an FFT on a time-domain averaged signal, 4096 points in length. Phase data were corrected for the phase angles of the primaries. The maximum level of the system distortion measured with an artificial ear cavity for the highest levels of primaries used in this study ($L_1=L_2=75$ dB SPL) was 70 dB below the primary level. Experimental control, data acquisition, and data analysis were performed using a PC with programs written in TESTPOINT (CEC, MA).

The following experimental procedures were used:

(i) DPOAE-grams (f_2 sweeps, f_2/f_1 ratio is constant, L_1 and L_2 are constant, L_2 is 10 dB below L_1) for low-level primaries were recorded on a regular basis during each experiment (up to 5 h) to confirm stable conditions of the animal. Data collected from an animal were rejected if DPOAE level changed by more than 5 dB at f_2 frequencies used for the other procedures.

(ii) DPOAE growth functions with increasing L_1 were measured while keeping L_2 and frequencies of primaries constant. The growth functions were measured at different f_2/f_1 ratios.

All procedures involving animals were performed in accordance with UK Home Office regulations.

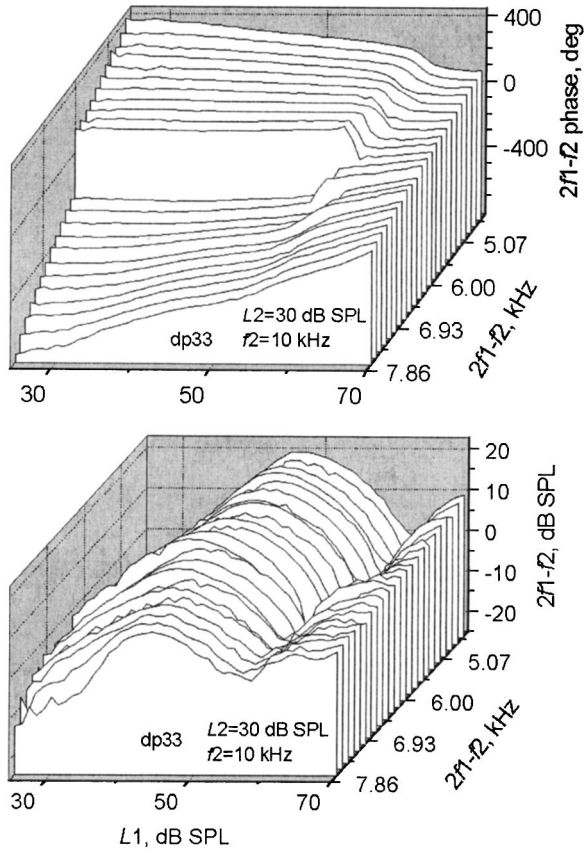


FIG. 1. Level dependence of the $2f_1-f_2$ DPOAE phase (top) and amplitude (bottom) at different DPOAE frequencies. L_2 and f_2 were held constant at 30 dB SPL and 10 kHz, respectively. f_1 was varied to obtain required $2f_1-f_2$ frequency.

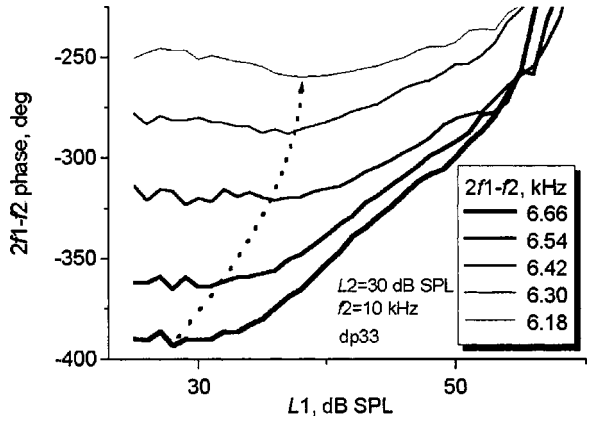


FIG. 2. Level dependence of the phase transition observed for the $2f_1-f_2$ DPOAE. High-resolution version of some of the curves shown in Fig. 1. The $2f_1-f_2$ DPOAE frequency for every curve is shown inside the panel. Dotted arrow indicates a shift of the transition between phase lag and phase lead observed with reduction of the $2f_1-f_2$ frequency.

III. RESULTS

The phase angle of the OHC mechanical response can change as a consequence of changes in the input to the OHCs (Kössl and Russell, 1992). It is possible to minimize this effect by using a specific experimental paradigm. Because the DPOAE is generated mainly from a region in the cochlea, which is close to the f_2 CF place (Brown and Kemp, 1984; Knight and Kemp, 2000), one can keep f_2 and L_2 constant and vary only the level L_1 of the low-frequency primary f_1. The frequency f_1 should be chosen to provide an f_2/f_1 ratio between 1.15 and 1.3. It has been reported (Kössl and Russell, 1992) that stimuli from this frequency range below the CF do not show level-dependent phase changes in the basal turn of the guinea pig cochlea at least up to the 70 dB SPL, the maximum level used in this study.

A. Frequency dependence of DPOAE phase changes

When DPOAEs are recorded during the experimental procedures defined above, then the phase of the $2f_1-f_2$ component leads with increasing L_1 for small f_2/f_1 ratios, i.e., when the $2f_1-f_2$ frequency is relatively high (Fig. 1). The phase lead gradually declines when f_1 moves further away from the high-frequency primary and the $2f_1-f_2$ frequency decreases accordingly. Eventually, the sign of the phase change inverses and the phase lags with increasing L_1 for low $2f_1-f_2$ frequencies. This sign inversion is expected if the $2f_1-f_2$ sweeps through the range of frequencies around a nonlinear resonance about a half of an octave below f_2. Then, the resonance frequency that corresponds to one of the frequencies of the normal modes for a given system coincides with the $2f_1-f_2$ value for which the phase angle is independent of the input signal level (L_1 in our case). Please note that the phase changes discussed in the current study are different from the relatively abrupt phase transition, which is observed for L_1 above 50 dB SPL and is most pronounced for the near resonance values of $2f_1-f_2$ (Fig. 1, top). The abrupt phase transition is associated with a notch in the DPOAE growth functions (Fig. 1, bottom). These transitions and corresponding notches are seen in the output of a nonlinear system at certain positions in the operating point of the nonlinearity (Frank and Kössl, 1996; Lukashkin and Russell, 1998, 1999, 2002; Bian *et al.*, 2002). These relatively fast phase transitions do not alter the general trend of the phase changes, i.e., whether the phase leads or lags in the region around the abrupt phase transition (Fig. 1, top). Essentially the same behavior of the DPOAE phase as that shown in Fig. 1 was observed in all 11 animals used in this study, and for f_2 ranged between 4 and 30 kHz.

The observed phase changes are not only frequency but also level dependent, which provides further evidence for the association of the phase transition with a nonlinear resonance. Namely, inversion of the direction of the phase changes is observed over a finite frequency range of DPOAE frequencies when f_1 is moved further away from f_2, so that the $2f_1-f_2$ frequency decreases (Figs. 1, 2). Initially only low-level responses are phase lagging, i.e., the curves in Fig. 2 have negative slopes at low levels, and the phase still leads at higher L_1. The phase lag spreads towards progressively higher L_1 with further reduction of the $2f_1-f_2$ frequency (dotted arrow in Fig. 2). This phenomenon is explained by the fact that the resonance frequency of a dynamic nonlinear system depends on the level of the input signals. The resonance frequency may shift up or down depending on the stimulus levels and on the properties of the particular nonlinear system. Accordingly, the phase pattern seen in Fig. 2 for the $2f_1-f_2$ component would be expected if the reso-

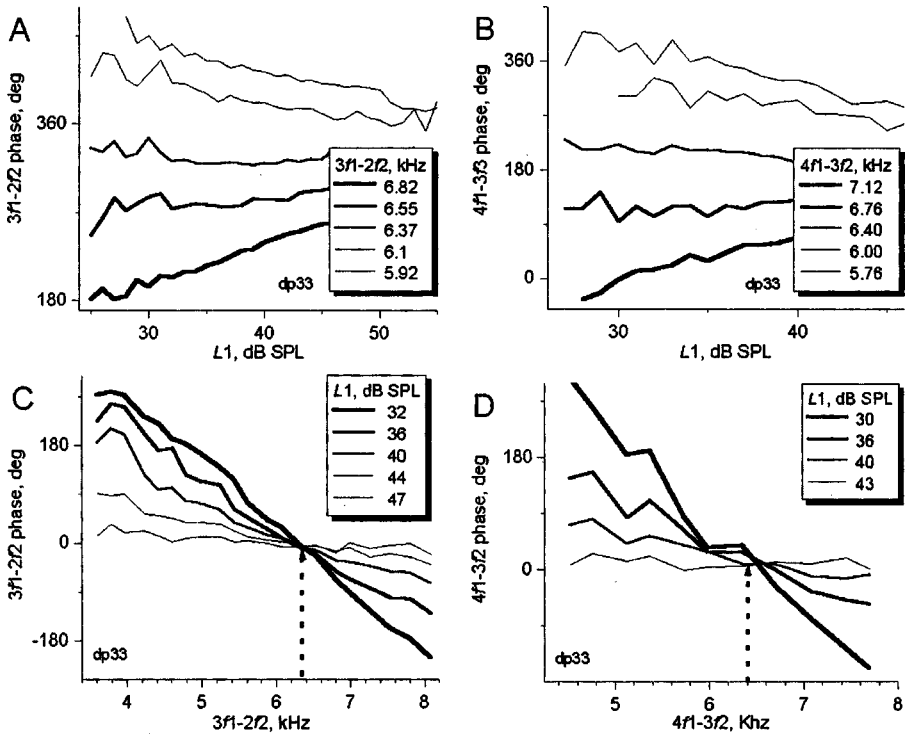


FIG. 3. Level dependence of the DPOAE phase for high-order components $3f_1-2f_2$ (left column) and $4f_1-3f_2$ (right column). L_2 and f_2 were held constant at 30 dB SPL and 10 kHz, respectively. f_1 was varied to obtain required DPOAE frequency [(A) and (B)]. The DPOAE frequency for every curve is shown inside the panels [(C) and (D)]. Phases expressed relative to the phases of emissions at $L_1=48$ dB SPL (C) and $L_1=44$ dB SPL (D). L_1 level for every curve is shown inside the panels. Dotted arrows indicate point of intersections, which correspond to the resonance frequency.

nance frequency becomes lower when the level of the stimulus (L_1) is increased.

B. Phase changes for different DPOAE frequency components

Two of the higher-order frequency components, $3f_1-2f_2$ and $4f_1-3f_2$, could be recorded reliably above the noise floor in our experiments. The dependence of their phase on L_1 [Figs. 3(A), (B)] is the same as that reported for the $2f_1-f_2$ component (Fig. 2); namely, level-dependent phase lead transforms into phase lag when the component's frequency falls below that of the postulated resonance. However, the phase transitions for the higher order DPOAEs occur within much narrower frequency ranges and the level dependence of the phase transition shown in Fig. 2 cannot usually be resolved for DPOAEs higher in order than $2f_1-f_2$. This abrupt change of the phase behavior is a consequence of the much steeper amplitude growth functions of the higher order frequency components at the output of a nonlinearity, a feature that is a general property of nonlinear systems. Figures 3(C), (D) show phase behavior for a much wider frequency range. Here, phase is normalized to its value at high L_1 levels. In this instance, zero value means that the phase does not change with increase in L_1, thereby indicating the resonance frequency. This frequency also coincides with the point of intersection of the phase plots for different L_1 levels. Theoretically, the resonance frequencies for all three low sideband DPOAEs found using the employed experimental paradigm should coincide, and this is indeed the case. These frequencies are almost identical for $3f_1-2f_2$ [~6.37 kHz, Figs. 3(A), (C)] and $4f_1-3f_2$ [~6.40 kHz, Figs. 3(B), (D)] components and they both fall within the frequency range (6.30–6.42 kHz), which can be determined for $2f_1-f_2$ component using different L_1 criteria (36–37 dB SPL) (Fig. 2). Some difference between resonance frequencies obtained using different DPOAE components might be anticipated because in our experiments we controlled only the sound levels but not the input to the DPOAE producing nonlinearity. However, the level of the input to the DPOAE producing nonlinearity at the DPOAE generation place can alter with changes in the f_2/f_1 ratio even when the levels of the primaries are kept constant (Lukashkin and Russell, 2001).

C. Map of the second resonance

The DPOAE is generated mainly from a location in the cochlea close to the f_2 CF place (Brown and Kemp, 1984; Knight and Kemp, 2000). Hence, it is possible to map the second resonance along the cochlea length by changing f_2 and finding the frequency of the transition, f_t, between the level-dependent phase lead and lag for every f_2 value. The $2f_1-f_2$ DPOAE component has much higher amplitude and can be reliably recorded above the noise floor for a wide range of L_1 [compare the horizontal axis in Fig. 2 and Figs. 3(A), (B)] Therefore, it is advantageous to use the $2f_1-f_2$ component to determine f_t. However, the transition occurs at different $2f_1-f_2$ frequencies for different values of L_1 (Fig. 2); hence, the map of the second resonance will be level specific. Figure 4 gives an example of f_t as a function of the high-frequency primary f_2 for one particular primary pair, $L_1=L_2+10$ dB$=40$ dB SPL. The phase transition takes place at lower $2f_1-f_2$ frequencies for the same L_2 but for a higher L_1 (Fig. 2). In this case the experimental points should be situated below the shown curve, reflecting a shift of the nonlinear resonance to lower frequencies. Correspondingly, a low L_1 criterion would cause an upward shift of the curve along the f_t axis, thereby reflecting an increase in the resonance frequency for lower levels of the input signals.

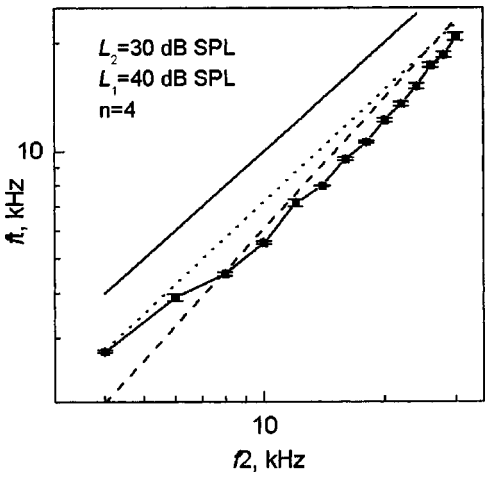


FIG. 4. Frequency of the second resonance as a function of f_2. Frequency of the phase transition, f_t, between level-dependent phase lag and lead is taken as a resonance frequency for particular levels of primary tones $L_1=40$ and $L_2=30$ dB SPL. Average for four animals. Solid straight line indicate dependence $f_t=f_2$. Earlier estimates (Allen and Fahey, 1993) of the second resonance map for cat ($f_t=0.08f_2^{1.22}$, dashed line) and human ($f_t=0.5f_2^{1.04}$, dotted line) are given for comparison.

It is worth noting that f_t is set at a frequency further below that of CF in the middle of a guinea pig's hearing range than it is at the extremes of the range (Fig. 4).

IV. DISCUSSION

We have shown that the phases of the DPOAEs change systematically with stimulus level for the DPOAE frequencies close to a frequency that is about a half octave below f_2. The phase changes are similar to those of BM responses that have been observed in the guinea pig for single tones in vicinity of the CF (Cooper and Rhode, 1992; Nuttall and Dolan, 1996) and indicate the presence of a second, nonlinear low-frequency resonance at each tonotopic place along the length of the cochlea. The resonance frequency is level dependent and shifts towards lower frequencies with increasing stimulus intensity. It is unlikely that there is a level-dependent inertia present in the cochlea. Therefore, the shift of the resonance to the lower frequencies indicates the presence of a nonlinear stiffness, which decreases with increasing intensity of stimulation (but see Shera, 2001). The existence of an intact TM is crucial for the presence of the low-frequency resonance; it disappears in mutant mice with a detached TM (Legan *et al.*, 2000). This last observation leads us to suggest that the TM forms a multiresonance complex with the OC. As a consequence, at each location along the length of the cochlea the complex has two normal modes of vibration. One mode is tuned to the frequency of the low-frequency resonance and the other is tuned to a higher frequency, which is close to the CF of the location.

The ability of the TM to oscillate at acoustic frequencies has been shown experimentally (Gummer *et al.*, 1996; Rhode and Cooper, 1996; Hemmert *et al.*, 2000) and incorporated into the realistic models of the cochlea (Zwislocki and Kletsky, 1979; Allen, 1980; Mammano and Nobili, 1993; Neely, 1993; Geisler and Sang, 1995; Markin and Hudspeth, 1995; Kolston, 1999). A model proposed by Allen and Fahey (1993) can qualitatively describe the data presented in this paper. Namely, the TM is elastically attached to the spiral limbus and can oscillate in the radial direction sliding over the reticular lamina (RL). The BM, loaded with the OC, forms a second resonator, which is coupled to the TM by the stiff OHC stereocilia. The system exhibits two resonances: a high-frequency resonance where the TM moves almost in counterphase with the RL and a resonance at lower frequencies when the TM and RL move almost in phase. The relative motion between the TM and the RL at the high-frequency resonance is optimal for stimulating both the OHCs and inner hair cells. There is little resultant shear displacement between the TM and RL during the low-frequency resonance. Consequently, excitation of the inner hair cell's stereocilia which are viscously coupled to the flow of fluid displaced by the OHC stereocilia, should be minimal. Accordingly, the thresholds of auditory afferent neurons that are tuned to the high-frequency resonance should be very high for stimuli around the low-frequency resonance. In this respect, it has been noted that the inner hair cells (Russell and Sellick, 1978) and the afferent fibers (Robertson, 1982; Salvi *et al.*, 1983; Liberman and Dodds, 1984) become sensitized for frequencies around the low-frequency resonance following surgical damage or selective depletion of the OHCs in the region of the CF. Under these conditions mechanical coupling between the TM and the OHCs would be expected to be greatly reduced, thereby also reducing synchrony in the movement between the TM and RL at the low-frequency resonance and increasing stimulation of the inner hair cells. Thus, in the intact cochlear, the low-frequency resonance improves the signal-to-noise ratio at the CF by isolating hair cells from excitation by low-frequency noise (Allen and Fahey, 1993). Shunting of the low-frequency excitation of the OHCs could also be important for stabilizing the operating point of the cochlear amplifier. The OHCs in the basal turn of the cochlea operate about the most sensitive region of their receptor-potential transfer function (Russell and Kössl, 1992), and even subtle low-frequency biasing of the operating point might lead to substantial reduction of the cochlear response. Earlier estimates of the low-frequency resonance (Allen and Fahey, 1993) showed its linear dependence on the CF in a log–log scale (Fig. 4). The current data (Fig. 4) reveal that in the region of the guinea pig's most sensitive hearing range (8–12 kHz) this resonance is set at a frequency further below that of the CF than it is at either end of the auditory range.

The input impedance of a system drops at the resonance frequency. Consequently, a secondary minimum in the BM iso-response-tuning curve may be recorded around the low-frequency resonance for measurements made in the hook region of the intact cochleae of the mouse (Legan *et al.*, 2000) and mustached bat (Russell and Kössl, 1999). Here, we present data that confirm the existence of the second resonance over a wide frequency range along the length the cochlea. The nonlinearity of both resonances, near the CF and below it, probably reflects the nonlinear pumping of energy into the cochlear partition from the OHCs. Augmentation of the BM mechanical response in the region of the tail of the frequency tuning curve following perfusion of the

cochlear with salicylate (Murugasu and Russell, 1995) also provides further evidence for the potential of the OHCs to influence BM mechanical responses at these frequencies. Accordingly, the augmentation of the response suggests that in the healthy cochlea, OHCs effectively suppress the BM movement for frequencies much below the CF.

It should be noted that the nonlinearity of the low-frequency resonance imposes limitations on techniques for analyzing DPOAE responses. For example, estimates of DPOAE group delays (Kemp and Brown, 1983) and DPOAE reconstructed filter functions (Stover *et al.*, 1996) have been employed to identify several sources of emission with different latencies. These techniques are based on studies of DPOAE phase changes during variation of the f_2/f_1 ratio. The effective level of the f_1 primary in the DPOAE generation place, which is close to the f_2 place (Brown and Kemp, 1984; Knight and Kemp, 2000), varies with variation of the f_2/f_1 ratio even if the primary SPLs are kept constant during this procedure. However the level dependent phase changes in the vicinity of a non-linear resonance are not associated with changes in time delays. Therefore techniques for linear system analysis should be used with great care for interpretation of the phase data obtained in experiments with variation of the f_2/f_1 ratio. Thus, the existence of a low-frequency mode of vibration at each frequency location in the mammalian cochlea and the presence of an associated nonlinear resonance of the system casts doubts on the validity of a number of techniques that have been used to date to analyze DPOAE behavior. The pattern of DPOAEs generated by the nonlinear cochlea provides a signature that can be decoded to reveal the functional properties of the cochlea. However, as this paper shows, in order to interpret this signature it is necessary to understand how the low-frequency resonance contributes in shaping the pattern of DPOAEs for wide range of stimulus parameters.

ACKNOWLEDGMENTS

This work was supported by a grant from the Wellcome Trust. We thank Guy Richardson for critical reading of an early version of the manuscript.

Allen, J. B. (**1980**). ''Cochlear micromechanics—A physical model of transduction,'' J. Acoust. Soc. Am. **68**, 1660–1670.

Allen, J. B., and Fahey, P. F. (**1993**). ''A second cochlear frequency map that correlates distortion product and neural tuning measurement,'' J. Acoust. Soc. Am. **94**, 809–817.

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.

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., and Kemp, D. T. (**1984**). ''Suppressibility of the $2f1-f2$ stimulated acoustic emissions in gerbil and man,'' Hear. Res. **13**, 29–37.

Cooper, N. P. (**1998**). ''Harmonic distortion on the basilar membrane in the basal turn of the guinea-pig cochlea,'' J. Physiol. (London) **509**, 277–288.

Cooper, N. P., and Rhode, W. S. (**1992**). ''Basilar membrane mechanics in the hook region of cat and guinea-pig cochleae: Sharp tuning and nonlinearity in the absence of baseline position shifts,'' Hear. Res. **63**, 163–190.

Cooper, N. P., and Rhode, W. S. (**1996**). ''Fast traveling waves, slow traveling waves, and their interactions in experimental studies of apical cochlear mechanics,'' Aud. Neurosci. **2**, 289–299.

Dallos, P., and Evans, B. N. (**1995**). ''High-frequency motility of outer hair cells and the cochlear amplifier,'' Science **267**, 2006–2009.

Davis, H. (**1958**). ''Transmission and transduction in the cochlea,'' Laryngoscope **68**, 359–382.

Frank, G., Hemmert, W., and Gummer, A. W. (**1999**). ''Limiting dynamics of high-frequency electromechanical transduction of outer hair cells,'' Proc. Natl. Acad. Sci. U.S.A. **96**, 4420–4425.

Frank, G., and Kössl, M. (**1996**). ''The acoustic two tone distortions $2f1-f2$ and $f2-f1$ and their possible relation to changes in the gain and the operating point of the cochlear amplifier,'' Hear. Res. **98**, 104–115.

Frolenkov, G. I., Belyantseva, I. A., Kurc, M., Mastroianni, M. A., and Kachar, B. (**1998**). ''Cochlear outer hair cell electromotility can provide force for both low and high intensity distortion product otoacoustic emissions,'' Hear. Res. **126**, 67–74.

Geisler, C. D., and Sang, C. N. (**1995**). ''A cochlear model using feed-forward outer-hair-cell forces,'' Hear. Res. **86**, 132–146.

Gummer, A. W., Hemmert, W., and Zenner, H. P. (**1996**). ''Resonant tectorial membrane motion in the inner ear: Its crucial role in frequency tuning,'' Proc. Natl. Acad. Sci. U.S.A. **93**, 8727–8732.

Hemmert, W., Zenner, H-P., and Gummer, A. W. (**2000**). ''Three-dimensional motion of the organ of Corti,'' Biophys. J. **78**, 2285–2297.

Kanis, L. J., and de Boer, E. (**1997**). ''Frequency dependence of acoustic distortion products in a locally active model of the cochlea,'' J. Acoust. Soc. Am. **101**, 1527–1531.

Kemp, D. T. (**1979**). ''Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea,'' Arch. Oto-Rhino-Laryngol. **224**, 37–45.

Kemp, D. T., and Brown, A. M. (**1983**). ''An integrated view of cochlear mechanical nonlinearities observable from the ear canal,'' in *Mechanics of Hearing*, edited by E. de Boer and M. A. Viergever (Delft University Press, Delft, The Netherlands), pp. 75–82.

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.

Kolston, P. J. (**1999**). ''Comparing *in vitro*, *in situ*, and *in vivo* experimental data in a three-dimensional model of mammalian cochlear mechanics,'' Proc. Natl. Acad. Sci. U.S.A. **96**, 3676–3681.

Kössl, M., and Russell, I. J. (**1992**). ''The phase and magnitude of hair cell receptor potentials and frequency tuning in the guinea pig cochlea,'' J. Neurosci. **12**, 1575–1586.

Legan, P. K., Lukashkina, V. A., Goodyear, R. J., Kössl, M., Russell, I. J., and Richardson, G. P. (**2000**). ''A targeted deletion in α-tectorin reveals that the tectorial membrane is required for the gain and timing of cochlear feedback,'' Neuron **28**, 273–285.

Liberman, M. C., and Dodds, L. W. (**1984**). ''Single neurone labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves,'' Hear. Res. **16**, 55–74.

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., and Russell, I. J. (**1998**). ''A descriptive model of the receptor potential nonlinearities generated by the hair cell mechanoelectrical transducer,'' J. Acoust. Soc. Am. **103**, 973–980.

Lukashkin, A. N., and Russell, I. J. (**1999**). ''Analysis of the f_2-f_1 and $2f_1-f_2$ distortion components generated by the hair cell mechanoelectrical transducer: Dependence on the amplitudes of the primaries and feedback gain,'' J. Acoust. Soc. Am. **106**, 2661–2668.

Lukashkin, A. N., and Russell, I. J. (**2001**). ''Origin of the bell-like dependence of the DPOAE amplitude on primary frequency ratio,'' J. Acoust. Soc. Am. **110**, 3097–3106.

Lukashkin, A. N., and Russell, I. J. (**2002**). ''Modifications of a single saturating nonlinearity account for post-onset changes in $2f_1-f_2$ distortion product otoacoustic emission,'' J. Acoust. Soc. Am. **112**, 1561–1568.

Mammano, F., and Nobili, R. (**1993**). ''Biophysics of the cochlea: Linear approximation,'' J. Acoust. Soc. Am. **93**, 3320–3332.

Markin, V. S., and Hudspeth, A. J. (**1995**). ''Modeling the active process of the cochlea: Phase relations, amplification, and spontaneous oscillation,'' Biophys. J. **69**, 138–147.

Murugasu, E., and Russell, I. J. (**1995**). ''Salicylate ototoxicity: The effects on basilar membrane displacement, cochlear microphonics, and neural responses in the basal turn of the guinea pig cochlea,'' Aud. Neurosci. **1**, 139–150.

Neely, S. T. (**1993**). ''A model of cochlea mechanics with outer hair cell motility,'' J. Acoust. Soc. Am. **94**, 137–146.

Nuttall, A. L., and Dolan, D. F. (**1996**). ''Steady-state sinusoidal velocity responses of the basilar membrane in guinea pig,'' J. Acoust. Soc. Am. **99**, 1556–1565.
Rhode, W. S., and Cooper, N. P. (**1996**). ''Nonlinear mechanics in the apical turn of the chinchilla cochlea *in vivo*,'' Aud. Neurosci. **3**, 101–121.
Robertson, D. (**1982**). ''Effects of acoustic trauma on stereocilia structure and spiral ganglion cell tuning properties in the guinea pig cochlea,'' Hear. Res. **7**, 55–74.
Robles, L., and Ruggero, M. A. (**2001**). ''Mechanics of the mammalian cochlea,'' Physiol. Rev. **81**, 1305–1352.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1991**). ''Two-tone distortion in the basilar membrane of the cochlea,'' Nature (London) **349**, 413–414.
Russell, I. J., and Kössl, M. (**1992**). ''Sensory transduction and frequency selectivity in the basal turn of the guinea pig cochlea,'' Philos. Trans. Soc., **336**, 317–324.
Russell, I. J., and Kössl, M. (**1999**). ''Micromechanical responses to tones in the auditory fovea of the greater mustached bat's cochlea,'' J. Neurophysiol. **82**, 676–686.
Russell, I. J., and Sellick, P. M. (**1978**). ''Intracellular studies of hair cells in the mammalian cochlea,'' J. Physiol. (London) **284**, 261–290.
Salvi, R. J., Hamernik, R. P., and Henderson, D. (**1983**). ''Response patterns of auditory nerve fibers during temporal threshold shift,'' Hear. Res. **10**, 37–67.
Shera, C. A. (**2001**). ''Intensity invariance of fine time structure in basilar-membrane click responses: Implications for cochlear mechanics,'' J. Acoust. Soc. Am. **110**, 332–348.
Stover, L. J., Neely, S. T., and Gorga, M. P. (**1996**). ''Latency and multiple sources of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **99**, 1016–1024.
Stover, L. J., Neely, S. T., and Gorga, M. P. (**1999**). ''Cochlear generation of intermodulation distortion revealed by DPOAE frequency functions in normal and impaired ears,'' J. Acoust. Soc. Am. **106**, 2669–2678.
van Dijk, P., and Manley, G. A. (**2001**). ''Distortion product otoacoustic emissions in the tree frog, *Hyla cinerea*,'' Hear. Res. **153**, 14–22.
Vetešník, A., and Nobili, R. (**2002**). ''Phase behaviour of the primaries in distortion product analysis,'' in *Biophysics of the Cochlea: from Molecules to Models* (Titisee, Germany), pp. 443–444.
Zheng, Y. L., Ohyama, K., Hozawa, K., Wada, H., and Takasaka, T. (**1997**). ''Effect of anesthetic agents and middle ear pressure application on distortion product otoacoustic emissions in the gerbil,'' Hear. Res. **112**, 167–174.
Zwislocki, J. J., and Kletsky, E. J. (**1979**). ''Tectorial membrane: A possible effect on frequency analysis in the cochlea,'' Science **204**, 639–641.

Beam and phase effects in angle-beam-through-transmission method of ultrasonic velocity measurement

L. Wang
Ohio State University, Nondestructive Evaluation Program, Edison Joining Technology Center, Columbus, Ohio 43221

A. I. Lavrentyev
Pratt & Whitney, East Hartford, Connecticut 06108

S. I. Rokhlin[a)]
Ohio State University, Nondestructive Evaluation Program, Edison Joining Technology Center, Columbus, Ohio 43221

(Received 10 June 2002; revised 18 October 2002; accepted 23 December 2002)

The accuracy of a plane wave approximation for phase velocity measurements in isotropic and anisotropic material using the angle-beam-through-transmission method has been investigated numerically and experimentally. In this method the velocity is measured in different propagation directions as a function of incidence angle. The effect of two factors on the measurement accuracy have been discussed: intrinsic phase shift of the transmitted signal through a fluid–solid interface and beam diffraction due to the finite beam size of receiver and transmitter. It is shown that the interface-induced phase shift can introduce an error in time delay measurements of the shear wave after the first critical angle and that this time delay error can be accurately corrected for. Numerical results obtained by a time-domain beam model show that except at the critical angles, the finite width of the transmitter and receiver only affects the amplitudes of the transmitted signals and has almost no effect on the measured zero-cross time delay; therefore the plane wave approximation for obtaining phase velocity from the measured time delay data by this method and the plane wave interface-induced phase correction are fully applicable. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1548151]

PACS numbers: 43.64.Kc, 43.58.Ry, 43.35.Cg [ASZ]

I. INTRODUCTION

Nondestructive determination of elastic constants of anisotropic materials, in particular composites, using ultrasonic phase or group velocity measurements has been an active field of study for the last several decades.[1–24] The elastic constants are obtained by inverting the Christoffel equation using a set of phase or group velocity data in different propagation directions. For an orthotropic material, seven of the nine elastic constants can be found[13–15] from measurements in two symmetry planes; the solution is independent of the initial guesses and is only slightly dependent on data scatter. The other two elastic constants can be found from measurements in nonsymmetry planes.[16] The nondestructive methods are different from early studies[3] produced on samples cut out in different directions relative to the material crystallographic system. In the nondestructive approach the samples are not cut out for the measurements; instead ultrasonic waves are excited and their velocity measured in different directions relative to the material axes. Different approaches to nondestructive measurement by through transmission and point source-point receiver methods have recently been reviewed in Refs. 23 and 24.

The accuracy of elastic constant determination depends on the precision of the velocity measurement. One method for phase velocity measurement, well-developed and simple to implement, is the self-reference double-through-transmission method.[13–21] If properly implemented this method can provide velocity measurements with precision 0.1–0.3%.[19,20] In this method, the sample is immersed in a fluid (water) and the ultrasonic signals are transmitted at different angles through the sample, reflected from the plane reflector and returned to the transducer after the second through transmission. The velocity in the sample is first measured at normal incidence while the velocity at oblique angles is determined from the signal time delay shift relative to that at normal incidence. The determination of velocity from the time delays uses a simple equation obtained for arbitrary anisotropy from plane wave geometrical acoustic analysis.[13–15] However, the phase velocity measurement based on these geometric considerations neglects two effects: one is the diffraction effect due to the limited size of the transducer, the other the signal distortion effect due to phase change at interfaces between dissimilar materials (for example, the interface between the test sample and the immersion fluid). The same situation occurs in different implementations of immersion methods for phase or group velocity measurement in through-transmission or reflection modes and also for angle beam velocity measurements using contact wedge transducers where the wedge material replaces the immersion fluid. The phase correction for the self-reference

[a)]Author to whom correspondence should be addressed. Electronic mail: rokhlin.2@osu.edu

double-through-transmission method has been briefly discussed by Lavrentyev and Rokhlin[17,19] and by Hollman and Fortunko,[20] who applied phase/slope analysis for the phase correction in this method. The literature on the diffraction correction of velocity measurements has a long history and has been reviewed recently.[25,26] The diffraction effect on velocity measurement is small [on the order of 0.1% (Ref. 25)], which is comparable with the error estimation for the self-reference double-through-transmission method.[19,20] However most previous analyses of the diffraction correction were performed for normally incident beams and isotropic materials. The beam effect on the velocity measurement at oblique incidence has not been investigated. It may be especially important in the vicinity of the critical angles where, as shown in this paper, the interface-induced intrinsic phase change of the transmitted signal is especially large and depends strongly on the incident angle. In addition when one performs angle beam measurements for anisotropic materials, such as composites, the dependence of the ultrasonic velocity on angle plays a significant role especially near the first critical angles when the velocity of the refracted longitudinal wave very strongly depends on the incident angle, thus introducing distortion in the transmitted ultrasonic pulse.

Assessment of plane wave approximation accuracy and phase correction are important when improved accuracy of angle beam velocity measurements is needed, for example for stress determination from velocity measurement.[19] In this paper, using a time-domain beam model we will numerically analyze these two effects on accuracy of angle beam velocity measurement for isotropic and anisotropic materials, discuss the phase correction and illustrate it by experimental examples. Examples are given for homogeneous isotropic or anisotropic (unidirectional graphite epoxy composite) materials; however, the conclusions are valid for measurements on multilayered structures[15,21] where for multidirectional composites one can consider Floquet waves[21] instead of bulk waves.

II. ULTRASONIC PHASE VELOCITY MEASUREMENT BY DOUBLE-THROUGH-TRANSMISSION SELF-REFERENCE METHOD AND PROBLEM STATEMENT

Let us first briefly consider the ultrasonic double-through-transmission measurement method described by Rokhlin *et al.*[13–16] The basic idea of this method is illustrated in Fig. 1(a). The sample is immersed in a fluid (water) and the ultrasonic signals are transmitted through the sample, reflected from the plane reflector and returned to the transducer after the second through transmission. First the velocity V_n in the sample and the time of flight t_n are measured for the double through transmission at normal incidence on the sample, then by continuously varying the incident angle, the time delay difference $\Delta t_{\theta_i}=t_{\theta_i}-t_n$ between the reference time of flight t_n at normal incidence and time of flight t_{θ_i} for an arbitrary oblique incident angle θ_i is measured (as we discuss below the time delay change is measured relative to the zero crossing between the signal maximum–minimum peak values). Considering the back reflector as a mirror, we

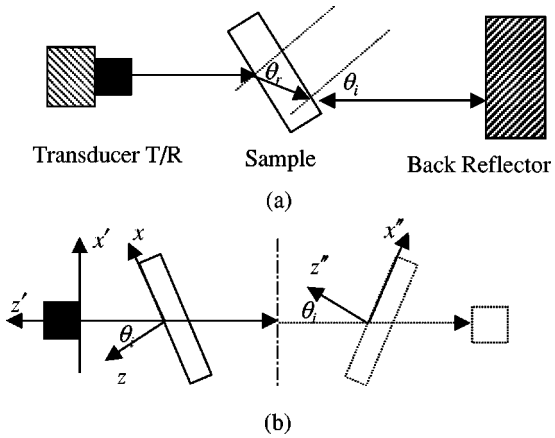


FIG. 1. (a) Schematic diagram of the self-reference bulk wave method. (b) Equivalent representation of the measurement used for modeling. The dashed lines indicate the mirror reflection of the back propagation path. Mirror plane is indicated by the broken vertical line.

represent in Fig. 1(b) the back propagation from the reflector to the transmitter/receiver by dotted lines on the right-hand side from the reflector (vertical midline). The coordinate systems associated with the transducers and the sample are used for modeling in Sec. IV.

The time delay Δt_{θ_i} at different propagation refraction angles θ_r corresponding to the incident angle θ_i shown in Fig. 1 for an arbitrary anisotropic material can be obtained from geometrical considerations[13–15]

$$\Delta t_{\theta_i}=\frac{h\cos(\theta_r-\theta_i)}{V_0\cos\theta_r}-\frac{2h}{1/V_0-1/V_n}-\frac{h}{V_p(\theta_r)\cos\theta_r},$$
$$\frac{\sin\theta_r}{V_p(\theta_r)}=\frac{\sin\theta_i}{V_0}, \tag{1}$$

where V_n is the velocity in the sample measured at normal incidence, V_0 is the sound speed in the immersion fluid, h is the thickness of the sample, and $V_p(\theta_r)$ is the phase velocity in the sample at angle θ_r, where θ_r is the refraction angle for the incident angle θ_i in the fluid. The phase velocity at normal incidence V_n is measured with high precision by overlapping multiply reflected signals from the front and back surface of the sample.

The phase velocity at refraction angle θ_r for a generally anisotropic material is calculated using the phase velocity in the normal direction V_n and the measured time-delay change Δt_{θ_i} with the rotation angle θ_i of the sample[14]

$$V_p(\theta_r)=\left(\frac{1}{V_n^2}+\frac{\Delta t_0-(\Delta t_0-\Delta t_{\theta_i})\cos\theta_i}{hV_0}-\frac{\Delta t_{\theta_i}(2\Delta t_0-\Delta t_{\theta_i})}{4h^2}\right)^{-1/2}, \tag{2}$$

with

$$\theta_r=\sin^{-1}\left(\frac{V_p(\theta_r)\sin\theta_i}{V_0}\right)$$

and

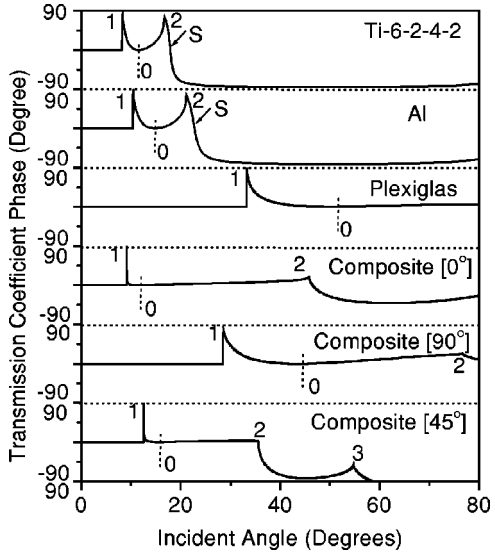


FIG. 2. Transmission phase shift at a water–solid interface for different materials. The longitudinal, fast shear and slow shear waves critical angles are marked by 1, 2, and 3, respectively. The zero phase angle is indicated by 0 and the surface wave critical angle is marked by *S*.

$$\Delta t_0 = 2h(1/V_0 - 1/V_n).$$

Since all measurements are made relative to the reference acoustic path in the presence of the sample, the effect of geometric imperfection is significantly reduced.[14] The elastic constants are determined from wave velocities in different propagation directions by inverting the Christoffel equation.[13,15,16]

Equations (1) and (2) for time delay and phase velocity calculations are based on geometric acoustics and do not account for the diffraction effect due to finite beam size and signal distortion due to phase change at interfaces. We will evaluate the applicability of these equations and their improvement by accounting for the phase correction.

III. PHASE CORRECTION IN PHASE VELOCITY MEASUREMENT

A. Phase change at fluid–solid interface

Reflection and transmission phenomena on the interface between fluid and isotropic or anisotropic solids have been well studied.[27–29] It has been shown that the reflection and transmission coefficients from the fluid/solid interface become complex when the incident angle is above the first critical angle and therefore the phase of the reflected or transmitted ultrasonic signal changes. Because for a semispace this phase shift is independent of frequency, the shape of the ultrasonic signal will be distorted, which may introduce error into the time delay measurement.

Figure 2 shows the computed phase shift for transverse wave transmission through a fluid–solid interface for different materials: Ti-6-2-4-2, aluminum, plexiglas and graphite/epoxy composite. The positions of longitudinal, fast shear and slow shear wave critical angles are marked by 1, 2, and 3, respectively. The position of the surface wave critical angle, where the phase equals zero, is marked by *S*. One can see that significant phase shift appears immediately above the first critical angle and before the second critical angle. The phase shift for the transmission coefficient in Ti-6-2-4-2 and Al has very sharp peaks at both critical angles. Plexiglas has significant phase shift at the second critical angle. The phase shift for a graphite/epoxy composite is small when the incident plane is oriented at 0° to the fiber direction and has similar behavior as plexiglas at 90° orientation.

Out of the plane of symmetry as shown in Fig. 2 for [45°] composite, three critical angles appear.[8] The behavior of the phase shift between the first and second critical angles is similar to that in isotropic materials, i.e., the phase shift jumps to 90° at the first critical angle, then decreases to zero at the zero phase angle (marked 0), defined below, and increases until the second critical angle. The phase shift is very small (<2°) between the first and second critical angles; however, after the second critical angle, the phase shift starts to decrease and reaches −90°. The phase shift in this range is significant and mainly negative.

As shown in Fig. 2, the phase shift for all materials has some common features: as the incident angle increases, the phase shift jumps from 0° to 90° at the first (longitudinal wave) critical angle, then decreases to 0° at a certain incident angle which we call the zero phase angle (ZPA, marked 0); finally above this angle the phase shift increases until the second (transverse wave) critical angle. After the second critical angle the phase shift decreases and passes through zero at the Rayleigh angle. In order to investigate the effect of material properties on the phase shift and the ZPA, it is advantageous to use analytical equations for the phase shift. For an orthotropic material in a plane of symmetry, the phase shift is zero before the first critical angle and for an incident angle θ_i above the first critical angle and below the second critical angle, it is given by

$$\tan\phi = \frac{T_t^x T_l^{z\prime} \cos\theta_i}{\rho_0 V_0^2 (W_l T_t^x - W_t T_l^x) - T_l^x T_t^z \cos\theta_i}, \tag{3}$$

where V_0 and ρ_0 are the velocity and density of the fluid; $W_{l,t}$ are normal displacement components in the solid corresponding to plane longitudinal (l) and transverse (t) waves, respectively, $T_{l,t}^z$ and $T_{l,t}^x$ are the corresponding components of normal and lateral stresses (see the Appendix) and $T_l^{z\prime}$ is the imaginary part of T_l^z. For an isotropic material, Eq. (3) simplifies to

$$\tan\phi = \frac{V_{l0}\cos^2(2\theta_t)\cos\theta_i}{(V_{l0}^2\sin^2\theta_i - 1)^{1/2}(V_{t0}^2\sin(2\theta_t)\sin 2\theta_i + \rho_0/\rho)}, \tag{4}$$

where $V_{l0} = V_l/V_0$ and $V_{t0} = V_t/V_0$; V_l, V_t are the longitudinal and transverse wave velocities in the solid, respectively; θ_t is the transverse wave refraction angle and θ_i is the incident angle in fluid.

At the longitudinal critical angle θ_c^I, one finds that the right-hand side of Eqs. (3) and (4) goes to infinity [$V_{l0}\sin\theta_i = 1$ in Eq. (4)], therefore the phase shift becomes 90°. At a certain incident angle above the first critical angle, the shear stress component T_t^x of the transverse wave in the solid equals zero and the right side of Eq. (3) becomes zero and thus the phase shift ϕ is zero. Physically this occurs at

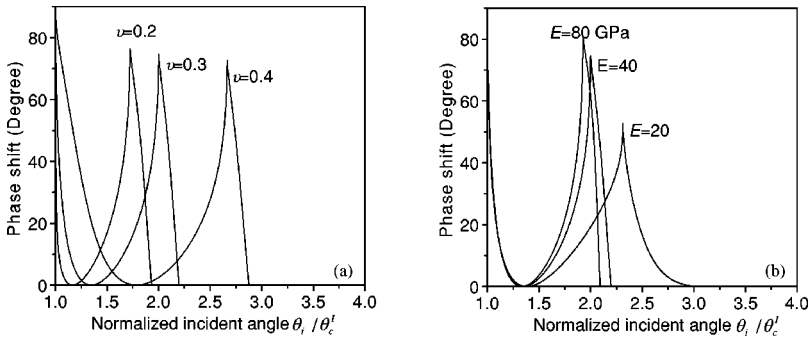


FIG. 3. Effect of material properties of solid on phase shift at a water–solid interface. Incident angle is normalized by the longitudinal critical angle. (a) Poisson ratio=0.2,0.3,0.4 (density=3.0 g/cm^3, Young's modulus=40 GPa). (b) Young's modulus=20,40,80 GPa (density=3.0 g/cm^3 and Poisson ratio=0.3).

ZPA incident angle to satisfy the boundary condition of total zero shear stress and therefore the vanishing of the shear stress component of the transverse wave in the solid requires zero shear stress produced at the interface by the refracted nonhomogeneous longitudinal wave. This condition can be satisfied only when the amplitude of the refracted nonhomogeneous longitudinal wave is zero, i.e., no refracted nonhomogeneous longitudinal wave occurs and only a pure transverse wave is excited in the solid. Because no nonhomogeneous wave is formed in the solid at the ZPA incident angle, the transmission coefficient is real and has zero phase. For an arbitrary isotropic solid, this condition is satisfied when the transverse wave is refracted at angle $\theta_t = 45°$ [$\cos(2\theta_t)=0$]. For an anisotropic solid this condition is not universal and the refracted angle at ZPA does not necessarily equal 45°. As the incident angle continuously increases above ZPA, the phase shift becomes nonzero again and at the transverse wave critical angle θ_c^{II}, the phase shift ϕ_c^{II} is given by

$$\tan\phi_c^{II} = \frac{\rho\cot(\theta_c^{II})}{\rho_f K^{1/2}}, \tag{5}$$

where $K=((C_{11}-C_{55})C_{33}-(C_{13}+C_{55})^2)/(C_{33}C_{55})$, C_{ij} are the material elastic constants. For isotropic solids,

$$K = 1 - \frac{V_t^2}{V_l^2} = \frac{1}{2(1-\upsilon)},$$

where υ is the Poisson ratio.

As discussed above, one can easily estimate the phase shift for different materials using the three special incident angles. The ZPA separates the phase shift into two regions. In the first region the phase decreases from 90° to 0° and in the second region, the phase increases from 0° to ϕ_c^{II}. In Fig. 3 we plot the phase shift for solids with different Young's modulus E and Poisson's ratio υ (the fluid is water). The incident angle is normalized by the longitudinal wave critical angle. As shown in Fig. 3, the normalized ZPA position is determined by the Poisson ratio; the higher the Poisson ratio, the larger the first region. Young's modulus E and density have very little effect on the first region. The phase shift in the second region is determined by the value of ϕ_c^{II} which is a function of Poisson ratio, Young's modulus and density. Smaller density and Young's modulus and larger Poisson ratio lead to smaller ϕ_c^{II}. For most materials, the Poisson ratio is around 0.3, therefore the angle range of the first region will be around half of the total angle range between the two critical angles. The phase shift at the second critical angle is usually higher for metals which have larger density and higher modulus. For plastics, due to their low density and modulus, the second critical angle is much larger and the phase shift is relatively small. For composites, one can see from Fig. 2 that the phase shift is similar to plexiglas at 90°orientation, while at 0°orientation, the ZPA occurs very close to the first critical angle and the phase change is small. The ϕ_c^{II} is small for a composite due to its low density.

The phase shift from solid to fluid exactly equals that from fluid to solid and is thus determined by Eq. (3) or (4). In the double-through-transmission measurement as shown in Fig. 1, the acoustic beam transmits twice from fluid to solid and from solid to fluid. Therefore the total phase shift in this measurement is four times the phase calculated by Eq. (3).

B. Signal distortion due to phase change

The plane wave phase shift at plane interfaces is frequency independent. For a harmonic signal $\sin(\omega t)$, introducing a phase shift ϕ is equivalent to shifting the signal in the time domain by $\Delta t=\phi/\omega$; i.e., $\sin(\omega t+\phi)=\sin(\omega(t+\Delta t))$. For a wideband signal, a constant phase shift to all harmonics may significantly change the shape of the signal. Here we will investigate this effect numerically for a typical Gaussian ultrasonic signal.

In the ultrasonic measurements, the transducer spectrum $F(\omega)$ can usually be approximated by a Gaussian function:

$$F(\omega) = \frac{1}{\sqrt{2\pi}B_w} e^{-(f-f_0)^2/2B_w^2}, \tag{6}$$

where f_0 is the center frequency and B_w is the half bandwidth. Then if we introduce a constant phase shift ϕ, the corresponding time domain signal is

$$f(t) = e^{-2\pi^2 t^2 B_w^2 + 2\pi i f_0(t-\phi/2\pi f_0)}. \tag{7}$$

From Eq. (7), one can see that for a Gaussian-shape signal, the phase shift effect is similar to that for a harmonic signal. When introducing a phase shift, the Gaussian envelope is not shifted and only the high frequency carrier filling the envelope shifts in time.

In order to simulate the phase shift effect on non-Gaussian ultrasonic signal propagation, we first measured an ultrasonic signal with a typical wide-band transducer with center frequency 5MHz shown in Fig. 4 (0° indicates no phase shift). Then we calculated the spectrum of this signal

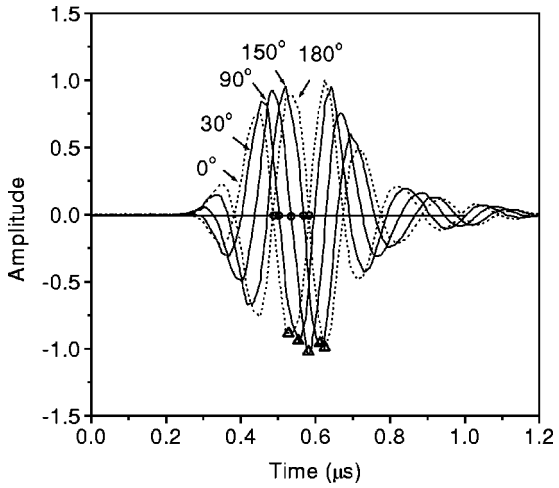


FIG. 4. Simulated distortion of the time-domain signal due to frequency-independent phase shift. The reference signal (no phase shift) is measured using a typical wide-band transducer with center frequency 5 MHz and bandwidth 2 MHz at −3 dB.

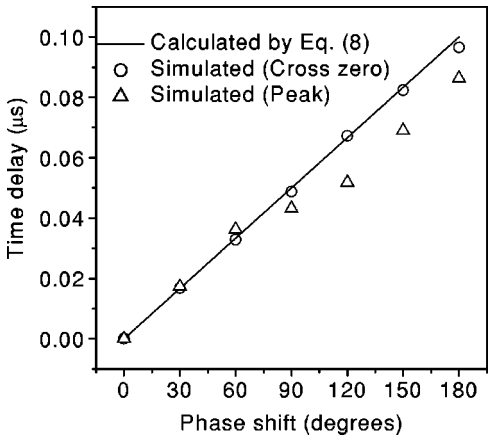


FIG. 5. Applicability of phase correction by Eq. (8) to time delay measurements using zero-cross and peak positions. Solid line is calculated from Eq. (8); circles and triangles are obtained from zero-crossing and peak positions of simulated signals (shown in Fig. 4), respectively.

using FFT and multiplied it by a constant phase to obtain the phase-shifted spectrum. The phase-shifted time domain signals shown in Fig. 4 are obtained using the inverse Fourier transform. When the phase shift is equal to 180 degrees, the obtained signal is the inversion of the original as shown by the two dotted lines in Fig. 4. While the peaks and zero-cross positions shift as the phase changes, the energy flow and the group velocity do not change.

C. Phase correction: simulation

Usually the time delay change is measured from the shift of a zero crossing or a peak position of the corresponding ultrasonic signal. As shown in Fig. 4, the phase shift changes these positions introducing an additional time delay. This effective time delay for a harmonic or Gaussian-shape signal can be calculated as

$$\Delta t_{\theta_i}^{P} = k\frac{\phi}{360f}, \tag{8}$$

where ϕ is the plane wave phase shift in degrees and f the ultrasonic wave frequency; the factor k is the number of fluid–solid, solid–fluid interfaces the wave is passing through: $k=4$ for the double through transmission method and $k=2$ in the single through transmission measurement. Equation (8) may be used to account for phase change at the interface and correct experimentally measured time delays (see the next section). In order to verify the applicability of the correction (8) for a typical experimental (non-Gaussian) ultrasonic signal shown in Fig. 4, we simulated the pulse transmitted through the interface and computed the time delays using the zero-cross positions and peak positions of the simulated signals as indicated by circles and triangles respectively in Fig. 4. Figure 5 compares the computed time delays from the time domain signals and those calculated from Eq. (8). It shows excellent accuracy of the correction (8) for predicting time delay change for the zero-cross measurements and less accuracy for the peak position measurements.

To illustrate the interface phase delay effect, we simulated the double-through-transmitted time domain signals through the solid layer in the fluid, and then obtained the time delay Δt_{θ_i} at different incident angles from these simulated signals. Figure 6(a) compares the time delay obtained from the simulated plane wave time domain signals transmitted through the 2.5 mm thick aluminum plate (circles) and the time delay calculated from Eq. (1) with (dashed lines) and without (solid lines) phase correction. For the shear wave, because its amplitude is very small for incident angle below the longitudinal critical angle, we only determined the time delay above the longitudinal critical angle. The center frequency of the ultrasonic signal is 5 MHz. One can see that the time delay calculated by Eq. (1) and obtained from the simulated time-domain signal agree well for the longitudinal wave (L) (before the first critical angle, standard deviation <0.001). However, after the first critical angle, the time delay of the shear wave (T) calculated by Eq. (1) and that obtained from the simulated signal delay have significant differences at angles close to the first and second critical angles. Due to the phase jump at the first critical angle as shown in Fig. 2, one can see this jump in the time delay measurement (indicated by an arrow). After phase correction the correspondence is excellent (standard deviation <0.0007). Figure 6(b) shows the results for a unidirectional composite lamina with thickness 2.5 mm at 0° orientation angle of the incident plane to the fiber direction. The properties of this lamina are given in Table I. As shown in Fig. 6(b), the time delay calculated by Eq. (1) without considering the phase correction is very close to that obtained from

TABLE I. Lamina properties. Axis 1 is in the plane of the plate and coincides with the fiber direction.

Lumina elastic constants (GPa)				
C_{11}	C_{22}	C_{55}	C_{12}	C_{23}
143.2	15.8	7.0	7.5	8.2
Density (g/cm^3)		Thickness (mm)		
1.6		2.5		

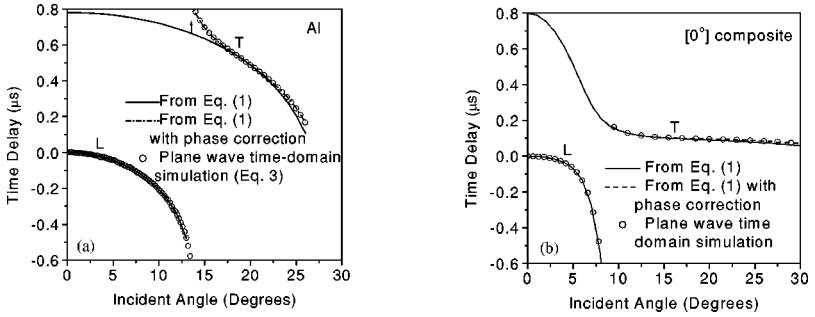


FIG. 6. Comparison of time of flight calculated with and without interface phase shift correction. Incident signal has center frequency 5 MHz and sample thickness is 2.5 mm. (a) Aluminum plate. (b) Unidirectional graphite/epoxy composite with incident plane orientation angle 0° to fiber direction.

the simulated time domain signals (standard deviation <0.0015) and phase correction may not be needed (compare to that for Al). Again, by considering the phase correction [Eq. (8)], the time delay determined by both methods coincides.

If the phase correction is not accounted for, then one introduces error into the measurement. One can estimate this error ε induced due to transmission through interfaces ($\Delta t^P_{\theta_i}$) normalizing by the time of flight t_n through the sample at normal incidence

$$\varepsilon = \frac{\Delta t^P_{\theta_i}}{t_n} = \frac{\phi V_n}{180 f h}. \tag{9}$$

Because the time of flight corresponding to shear wave propagation at oblique incidence through the sample is usually larger than the time of flight t_n at normal incidence, the phase-induced error estimated by Eq. (9) is the upper bound of the actual phase-induced error at arbitrary incident angle. For example, for a 2.5 mm aluminum, the upper bound of the error in double-through-transmission measurement at 5 MHz is 2.8% for a 20° phase shift.

D. Phase correction: experiment

To calculate the phase change at the interface one needs to know the elastic properties of the solid, which are a priori unknown before measurement. To correct the measured time delay data by taking into account the phase shift at the interface, we use the following procedure. At first the wave velocity is calculated using (2) from the measured time delays without phase correction and the elastic constants of the sample are determined in the first approximation. Based on the elastic constants found, the shear wave phase shift is calculated using Eq. (3). The calculated phase shift is used to correct the experimentally measured time delay data (corrected time delay equals directly measured time delay plus the time delay $\Delta t^P_{\theta_i}$ introduced by the phase shift) and to calculate the phase velocity again using Eq. (2). The procedure can be repeated to improve the precision of the velocity measurement using iterations; however the first iteration is usually satisfactory. Finally, the sample elastic constants are determined from the corrected time delay data.

An experimental example for the shear wave velocity phase correction is shown in Fig. 7 for (a) plexiglas (9.5 mm) and (b) Ti-6-2-4-2 samples (3 mm). Due to the large phase shift just after the first critical angle for plexiglas (shear wave refraction angle around 30°), one can observe a sharp dip in the velocity obtained before performing the phase correction. For Ti-6-2-4-2, the phase shift is significant around both the first (shear wave refraction angle 30°) and second (shear wave refraction angle close to 90°) critical angles. For these samples and frequencies of measurements, the difference between phase velocity measurements with and without phase correction can be significant. For lower frequency and thinner samples, the difference could be even higher.

Hollman and Fortunko[20] described a different phase correction method. It is based on phase/slope analysis[30,31] of the phase spectrum of the transmitted signal. The interface-induced phase change is frequency independent and is separated from the frequency dependent part of the phase spectrum which is related to the signal time of flight. This method provides a good alternative to the phase correction method described above; however it is more difficult to implement.

The relative error in reconstructed elastic constants from velocity data is much smaller than that of the maximum velocity error due to interface-induced phase delay. This is the result of the least squares minimization at all measured angles[13,15,16] and therefore the deviation of the velocity from the actual velocity in a small range of angle does not produce a drastic effect on the elastic constants determined. However the phase correction is very simple to implement and is recommended for all measurements.

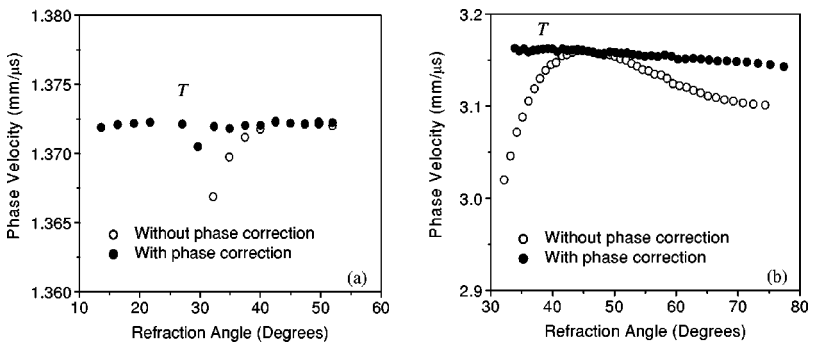


FIG. 7. Measured angular dependence of transverse wave phase velocity with and without phase correction. (a) Plexiglas, (b) Ti-6-2-4-2 plate obtained by the plasma sprayed process.

IV. BEAM EFFECT ON PHASE VELOCITY MEASUREMENT

A. Time domain beam model

It was shown above that the interface-induced phase shift significantly affects time delay measurements at incident angles close to the critical angles. This phase shift depends strongly on the incident angle. Rapid change of transmission amplitude and phase leads to distortion of the transmitted signal and since the signal from a finite size source is represented by a spectrum of incident angles, it is important to evaluate the beam effect on the velocity measurements in this experimental arrangement, and specifically to demonstrate the applicability of the plane wave phase correction described above. Also the strong dependence of longitudinal velocity on propagation angles in composites introduces additional signal distortion.

To simulate double-through-transmission measurements, we will use a two-dimensional time-domain beam model based on a wave number integration technique.[29] We apply it to the experimental arrangement of the double-through-transmission method shown in Fig. 1(b) considering the back reflector as a mirror. Accounting for the phase delay in fluid as shown in Fig. 1(b), one can obtain the time-domain voltage output for the 2D approximation:

$$V_{\text{out}}(t)=\int_{-\infty}^{+\infty}F(\omega)e^{i\omega t}d\omega\int_{-\infty}^{+\infty}\Phi_t(k_x')T_t(k_x)T_b(k_x'')\times e^{i(-2k_z'L+k_zd+k_z''d)}\,dk_x', \tag{10}$$

where L is the distance between the surfaces of the transducer and back reflector, d is the thickness of the sample. $k_z=\sqrt{(\omega/V_0)^2-k_x^2}$, V_0 is the acoustic velocity in the fluid. k_x',k_z' and k_x'',k_z'' are the wave numbers in the coordinate systems rotated from the axis z [Fig. 1(b)] by θ_i (primed) and $2\theta_i$ (double primed), respectively (z' is the normal to the transducer face). $T_t(k_x)$ and $T_b(k_x'')$ are the transmission coefficients through the layer for the wave incident from fluid on the layer top or bottom respectively. $F(\omega)$ is the frequency response of the transmitter/receiver; $\Phi_t(k_x')$ is the angular response of the transducer. We use Gaussian functions to approximate the angular responses of the transmitter/receiver

$$\Phi_t(k_x')=\frac{B}{\sqrt{2\pi}}e^{-(k_x'B)^2/2}, \tag{11}$$

where B is the beam width of the transmitter or receiver. The transducer frequency response $F(\omega)$ and B in Eq. (11) are experimentally determined by optimized fitting of the measured reflection signal from a homogeneous semispace such as a block of aluminum. The integration bound in Eq. (10) is determined by $\Phi_t(k_x')\leqslant\varepsilon$ where ε is a very small number; we used $\varepsilon=0.001$. In our double-through-transmission measurements we set the sample in the far field; therefore the nonhomogeneous waves are not received by the transducer and one need integrate only over the k_x' interval when the corresponding $k_z'=\sqrt{(\omega/V_0)^2-(k_x')^2}$ is real. For the plane wave approximation (Sec. II C), the transducer angular response is represented by a delta function: $\Phi_t(k_x')=\delta(k_x')$.

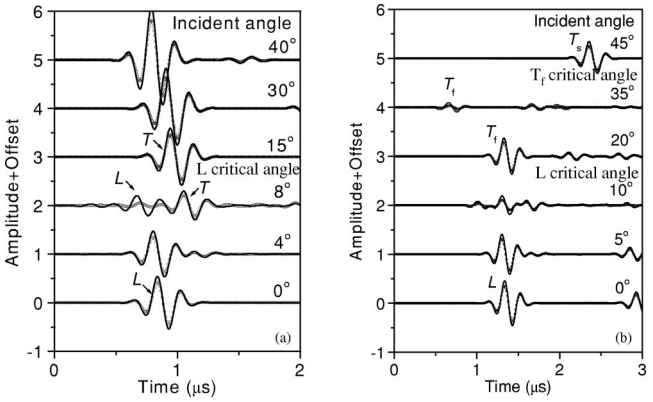


FIG. 8. Comparison of time domain double transmitted signal calculated by beam (open circles) and plane wave (solid lines) models at different incident angles on unidirectional graphite/epoxy sample. Only the first group of transmitted signals is shown. Beam width B is 5 mm and distance between transducer and reflector L is 80 cm. Incident signal has center frequency 5 MHz. (a) incident plane orientation to fiber direction is 0°, (b) incident plane orientation angle is 45°. T_f labels the fast quasitransverse wave, T_s the slow quasitransverse wave.

B. Beam effect on phase velocity measurement

In Sec. II using the plane wave model we investigated the accuracy of Eq. (2) for phase velocity measurement; here we will estimate the finite beam effect in these measurements. Figure 8 compares the ultrasonic signals obtained using the beam model [Eq. (10)] (open circles) and the plane wave approximation (solid lines) for a unidirectional graphite/epoxy composite at 0° (a) and 45° (b) incident plane orientations to the fiber direction. Composite properties are given in Table I. In the beam model calculations, we use beam width $B=5$ mm. At small incident angles, one observes only the longitudinal wave (marked as L). The shear waves start to appear when the incident angle is larger than 5° (marked by T). Behind the first transmitted signal one can also observe multiple reflections in the layer. For phase velocity measurements [Eq. (2)] we use only the first double-through-transmitted signal. As shown in Fig. 8 for different incident angles, only the signal amplitude has been affected by the finite width of the transmitter. It has almost no effect on the zero crossing time delay. The signals calculated by both models have the same shape, except near the critical angle (8°). At incident angles close to the critical angle, due to the sharp phase shift the refracted ultrasonic beam distorts and the plane wave approximation becomes invalid. In practice one should avoid phase velocity measurement when observing strong shape variation of the ultrasonic signal near the critical angle.

The beam effect on time delay measurements is shown in Fig. 9, where we replot Fig. 6 adding the time delay obtained from the beam time domain model (indicated by crosses). As one can see from Fig. 9(b), some discrepancy between the plane wave and beam models occurs for longitudinal waves immediately below the first critical angle. Since the interface-induced phase shift for the refracted longitudinal wave is zero, and the discrepancy for aluminum is

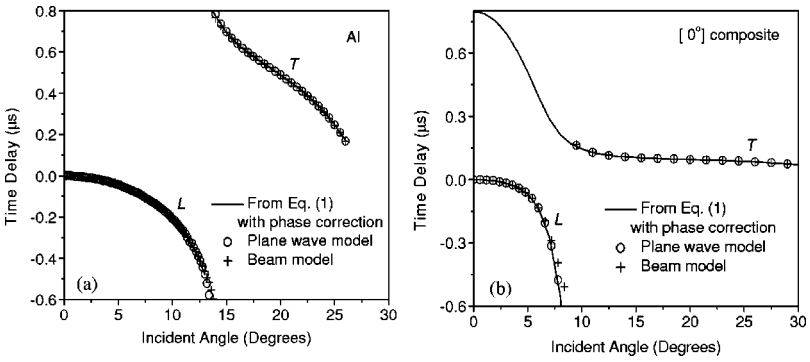


FIG. 9. Comparison of time of flight calculated using plane and beam models and by Eq. (1) with and without (below first critical angle) phase correction. Measurement conditions are the same as those in Fig. 6. (a) Aluminum plate. (b) Unidirectional graphite/epoxy composite with incident plane orientation angle 0° to fiber direction.

negligible [Fig. 9(a)], one can attribute the differences between the two models [Fig. 9(b)] to the strong anisotropy of the composite. However, one should note that in the experiment such high refraction angles for the longitudinal wave are usually unattainable.[13,14] However in general, as can be seen (Fig. 9), the discrepancy in time delays for transverse wave calculated by Eq. (1) with phase correction by plane wave model and by beam model is very small (standard deviation $<0.2\%$). This conclusion is similar to those obtained in the literature[25,26] for other experimental arrangements.

V. CONCLUSION

The double-through-transmission self-reference method has been analyzed using a time-domain beam model. It is shown theoretically and experimentally that the phase shift at the fluid–solid interface should be corrected for most materials for incident angles above the first critical angle. This phase shift effect decreases with frequency and sample thickness. It was demonstrated that for an ultrasonic signal with Gaussian shape this phase shift effect can be very accurately corrected by a phase shift calculated from the plane transverse wave transmission coefficient. It is also shown that for accurate results a zero crossing in the middle of the signal should be used for time delay measurements. The beam diffraction effect has been investigated theoretically. It is shown that in the far field the finite width of the transmitter has an effect under 0.2% on the time delay measurement in the double-through-transmission method and affects only the amplitude of the transmitted signal. Therefore the plane wave phase correction method developed in this paper is sufficient.

APPENDIX: PHASE SHIFT AT FLUID–SOLID INTERFACE

For incident angle θ_i above the first (longitudinal wave) critical angle, the k_z^l component of the quasilongitudinal wave number becomes complex and the quasitransverse k_z^t is real. The normalized wave numbers are given by

$$\alpha_l=\sqrt{[-B-(B^2-4AC)^{1/2}]/2A},$$

and

$$\alpha_t=\sqrt{[-B+(B^2-4AC)^{1/2}]/2A}, \tag{A1}$$

where

$$A=C_{33}C_{55},$$

$$B=(C_{11}\sin^2\theta_i-\rho V_0^2)C_{33}+(C_{55}\sin^2\theta_i-\rho V_0^2)C_{55}-(C_{13}+C_{55})^2\sin^2\theta_i, \tag{A2}$$

$$C=(C_{11}\sin^2\theta_i-\rho V_0^2)(C_{55}\sin^2\theta_i-\rho V_0^2).$$

where $\alpha_{l,t}=k_z^{l,t}/k_0$, k_0 is the wave number in water. C_{ij} are the elastic constants. ρ is the density of the solid. The normal $W_{l,t}$ and lateral $U_{l,t}$ displacement components corresponding to a plane longitudinal (l) and transverse (t) wave in the solid substrate are given by

$$W_{l,t}=C_{11}\sin^2\theta_i+C_{55}(\alpha_{l,t})^2-\rho V_0^2,$$
$$U_{l,t}=(C_{13}+C_{55})\alpha_{l,t}\sin\theta_i. \tag{A3}$$

The normal $T_{l,t}^z$ and shear stresses $T_{l,t}^x$ are given by

$$T_{l,t}^z=(C_{13}U_{l,t}\sin\theta_i+C_{33}\alpha_{l,t}W_{l,t}),$$
$$T_{l,t}^x=C_{55}(\alpha_{l,t}U_{l,t}+W_{l,t}\sin\theta_i). \tag{A4}$$

Here the stresses are normalized by ik_0. The transverse wave transmission coefficient is calculated using the boundary conditions at the fluid–solid interface. From the complex transmission coefficient, one can obtain the phase which is given in Eq. (3).

[1] M. F. Markham, "Measurement of the elastic constants of fiber composites by ultrasonics," Composites **1**, 145–149 (1970).
[2] R. E. Smith, "Ultrasonic elastic constants of carbon fibers and their composites," J. Appl. Phys. **43**, 2555–2562 (1974).
[3] R. D. Kriz and W. W. Stinchomb, "Elastic moduli of transversely isotropic graphite fibers and their composite," Exp. Mech. **19**, 41–49 (1979).
[4] B. Hosten, M. Deschamps, and B. R. Tittman, "Inhomogeneous wave generation and propagation in lossy anisotropic solids. Application to the characterization of viscoelastic composite materials," J. Acoust. Soc. Am. **82**, 1763–1770 (1987).
[5] L. H. Pearson and W. J. Murri, "Measurment of ultrasonic wavespeeds in off-axis directions of composite materials," in *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 6B, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1987), pp. 1093–1101.
[6] S. I. Rokhlin and W. Wang, "Ultrasonic evaluation of in-plane and out-plane elastic properties of composite materials," in *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 8B, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1989), pp. 1489–1496.
[7] W. R. Rose, S. I. Rokhlin, and L. Adler, "Evaluation of anisotropic properties of graphite-epoxy composite plate using Lamb waves" in *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 6B, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1987), pp. 1111–1118.
[8] S. I. Rokhlin and W. Wang, "Critical angle measurement of elastic costants in composite material," J. Acoust. Soc. Am. **86**, 1876–1882 (1989).
[9] A. G. Every and W. Sachse, "Determination of the elastic constants of anisotropic solids from acoustic-wave group-velocity measurements," Phys. Rev. B **42**, 8196–8205 (1990).

[10] B. Hosten, ''Stiffness matrix invariants to validate the characterization of composite materials with ultrasonic methods,'' Ultrasonics **30**, 365–371 (1992).
[11] K. Y. Kim, T. Ohtant, A. R. Baker, and W. Sachse, ''Determination of all elastic-constants of orthotropic plate specimens from group-velocity data,'' Res. Nondestruct. Eval. **7**, 13–29 (1995).
[12] A. D. Degtyar and S. I. Rokhlin, ''Comparison of elastic constants determination in anisotropic material form ultrasonic group and phase velocity data,'' J. Acoust. Soc. Am. **102**, 3458–3466 (1997).
[13] S. I. Rokhlin and W. Wang, ''Double through-transmission bulk wave method for ultrasonic phase velocity measurement and determination of elastic constants of composite materials,'' J. Acoust. Soc. Am. **91**, 3303–3312 (1992).
[14] Y. C. Chu and S. I. Rokhlin, ''Comparative analysis of through-transmission ultrasonic bulk wave methods for phase velocity measurements in anisotropic materials,'' J. Acoust. Soc. Am. **95**, 3204–3212 (1994).
[15] Y. C. Chu and S. I. Rokhlin, ''A method for determination of elastic constants of a unidirectional lamina from ultrasonic bulk velocity measurements on [0/90] cross-ply composites,'' J. Acoust. Soc. Am. **96**, 342–354 (1994).
[16] Y. C. Chu, A. D. Degtyar, and S. I. Rokhlin, ''On determination of orthotropic material moduli from ultrasonic velocity data in nonsymmetry planes,'' J. Acoust. Soc. Am. **95**, 3191–3203 (1994).
[17] A. I. Lavrentyev and S. I. Rokhlin, ''Phase correction for ultrasonic bulk wave measurements of elastic constants in anisotropic materials,'' in *Review of Progress in QNDE,* Vol. 16B, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1997), pp. 1367–1374.
[18] K. Kawashima, I. Fujii, A. Yamamoto, Y. Shimizu, and A. Fukushima, ''Measurement of anisotropic elastic constants of sandwiched SiC/Ti composite by double transmission, critical angle and leaky surface wave techniques,'' in *Review of Progress in Quantitative Nondestructive Evaluation,* Vol. 16B, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1997), pp. 1135–1142.
[19] A. I. Lavrentyev, A. D. Degtyar, and S. I. Rokhlin, ''New method for determination of applied and residual stresses in anisotropic material from ultrasonic velocity measurement,'' Mater. Eval. **55**, 1162–1168 (1997).
[20] K. W. Hallman and C. M. Fortunko, ''An accurate method for measurement of transverse elastic-wave velocities,'' Meas. Sci. Technol. **9**, 1721–1727 (1998).
[21] L. Wang and S. I. Rokhlin, ''Floquet wave ultrasonic method for determination of single ply moduli in multi-directional composites,'' J. Acoust. Soc. Am. **112**, 916–924 (2002).
[22] C. Aristegui and S. Baste, ''Determination of the elastic symmetry of a monolithic ceramic using bulk acoustic waves,'' J. Nondestruct. Eval. **19**, 115–127 (2000).
[23] B. Hosten, ''Ultrasonic through-transmission method for measuring the complex stiffness moduli of composite materials,'' in *Handbook of Elastic Properties of Solids, Liquids, and Gases,* edited by M. Levy, H. E. Bass, and R. R. Stern (Academic, New York, 2001), Vol. I, Chap. 3, pp. 67–85.
[24] A. E. Every, K. Y. Kim, and W. Sachse, ''Point source/point-receiver method,'' in *Handbook of Elastic Properties of Solids, Liquids, and Gases,* edited by M. Levy, H. E. Bass, and R. R. Stern (Academic, New York, 2001), Vol. I, Chap. 4, pp. 87–106.
[25] A. Ruiz and P. B. Nagy, ''Diffraction correction for precision surface acoustic wave velocity measurements,'' J. Acoust. Soc. Am. **112**, 835–842 (2002).
[26] J. Kushibiki and M. Arakawa, ''Diffraction effects on bulk-wave ultrasonic velocity and attenuation measurements,'' J. Acoust. Soc. Am. **108**, 564–573 (2000).
[27] B. A. Auld, *Acoustic Fields and Waves in Solids* (Krieger, Malabar, Florida, 1990).
[28] A. H. Nayfeh, *Wave Propagation in Layered Anisotropic Media* (North-Holland, Amsterdam, 1995).
[29] L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1960).
[30] W. Sachse and Y. H. Pao, ''On determination of phase and group velocities of dispersive waves in solids,'' J. Appl. Phys. **49**, 4320–4327 (1978).
[31] S. I. Rokhlin, D. K. Lewis, K. F. Graff and L. Adler, ''Real-time study of frequency dependence of attenuation and velocity of ultrasonic waves during the curing reaction of expoxy resin,'' J. Acoust. Soc. Am. **79**, 1786–1793 (1986).

Effects of peripheral nonlinearity on psychometric functions for forward-masked tones[a)]

Kim S. Schairer,[b)] Lance Nizami, Jason F. Reimer,[c)] and Walt Jesteadt
Center for Hearing Research, Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 14 March 2002; revised 8 October 2002; accepted 16 December 2002)

Psychometric functions (PFs) for forward-masked tones were obtained for conditions in which signal level was varied to estimate threshold at several masker levels (variable-signal condition), and in which masker level was varied to estimate threshold at several signal levels (variable-masker condition). The changes in PF slope across combinations of masker frequency, masker level, and signal delay were explored in three experiments. In experiment 1, a 2-kHz, 10-ms tone was masked by a 50, 70 or 90 dB SPL, 20-ms on-frequency forward masker, with signal delays of 2, 20, or 40 ms, in a variable-signal condition. PF slopes decreased in conditions where signal threshold was high. In experiments 2 and 3, the signal was a 4-kHz, 10-ms tone, and the masker was either a 4- or 2.4-kHz, 200-ms tone. In experiment 2, on-frequency maskers were presented at 30 to 90 dB SPL in 10-dB steps and off-frequency maskers were presented at 60 to 90 dB SPL in 10-dB steps, with signal delays of 0, 10, or 30 ms, in a variable-signal condition. PF slopes decreased as signal level increased, and this trend was similar for on- and off-frequency maskers. In experiment 3, variable-masker conditions with on- and off-frequency maskers and 0-ms signal delay were presented. In general, the results were consistent with the hypothesis that peripheral nonlinearity is reflected in the PF slopes. The data also indicate that masker level plays a role independent of signal level, an effect that could be accounted for by assuming greater internal noise at higher stimulus levels. [DOI: 10.1121/1.1543933]

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

I. INTRODUCTION

Forward masking occurs when the threshold for detection of a brief tone or signal is elevated in the presence of a preceding stimulus or masker. Forward masking data have been obtained in conditions in which signal level is varied to estimate threshold at fixed masker levels (*variable-signal condition*), or conditions in which masker level is varied to estimate threshold at fixed signal levels (*variable-masker condition*). The function describing the relation between masker level and signal level at threshold is referred to as growth of masking (GOM). These patterns have been described in detail for variable-signal (e.g., Jesteadt *et al.*, 1982) and variable-masker conditions (e.g., Nelson and Freyman, 1987), and will not be reviewed here. A recent model of forward masking (Oxenham and Plack, 2000; Plack and Oxenham, 1998) accounts for GOM results by assuming a peripheral nonlinearity with a specific form that is consistent with direct basilar membrane (BM) input–output (I/O) recordings (Rhode, 1971; Robles *et al.*, 1986; Ruggero *et al.*, 1997) and with estimates from auditory nerve rate-intensity functions (Yates *et al.*, 1990). The BM I/O function is described as having a compressive, nonlinear growth when the signal frequency is equal to the characteristic frequency (CF), which is analogous to the case in which the masker and signal are the same frequency (on-frequency masker), and more linear growth when the signal frequency is different from CF, which is analogous to the case in which the masker is a different frequency than the signal (off-frequency masker). The nonlinearity at CF can be characterized by two straight lines in log–log coordinates, with a low-level linear region (i.e., slope of 1.0), and a moderate-level compressive region with a slope of around 0.2 above some breakpoint (Yates *et al.*, 1990).

The Plack and Oxenham (1998) model assumes that a sliding temporal integration window receives the output of the peripheral nonlinearity, and serves to overlap the masker and signal. In a two-interval forced choice task (2IFC), the output of the temporal window is compared for the two intervals, and the subject is assumed to vote for the interval with the greater output. Multiplicative internal noise limits performance and is assumed to be uniform across all conditions. The current paper is concerned only with the peripheral nonlinearity component of the model. Therefore, the remaining components of the model and their underlying processes will not be discussed further here.

The Plack and Oxenham (1998) model describes the relationship between signal level and masker level when both of these quantities are changing, as in GOM functions. The

[a)]Portions of this work were presented in ''Psychometric functions for detection of tones following a forward masker,'' 2001 MidWinter Meeting of the Association for Research in Otolaryngology, St. Petersburg Beach, FL, February 2001; ''Psychometric functions for detection of tones following on- and off-frequency forward maskers,'' 141st Meeting of the Acoustical Society of America, Chicago, IL, June 2001; and ''Psychometric functions for variable masker levels in the context of peripheral nonlinearity,'' 2002 MidWinter Meeting of the Association for Research in Otolaryngology, St. Petersburg Beach, FL, January 2002.

[b)]Electronic mail: schairerk@boystown.org
[c)]Current address: Department of Psychology, California State University, San Bernardino, 5500 University Parkway, San Bernardino, CA 92407-2397.

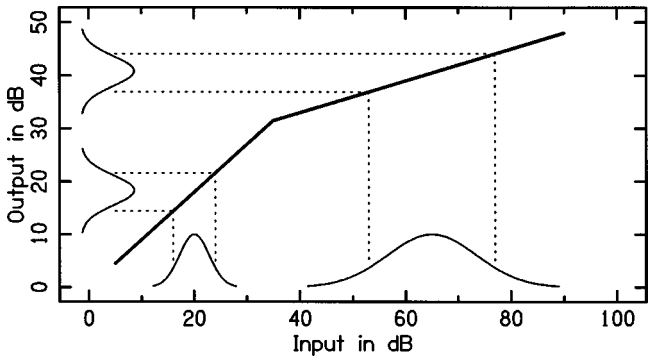


FIG. 1. Example of a compressive nonlinearity similar to the nonlinearity assumed in the model of forward masking. For conditions that produce low signal level at masked threshold, the signal passes through the more linear portion of the function as its level varies during the threshold estimation procedure. Changes in signal level at the input of the function will result in similar changes in signal level at the output of the function, and a smaller range of input levels will be presented in establishing threshold. The underlying distribution of input signal levels in this case is represented by the probability density function (PDF) on the left along the *X* axis, and the corresponding distribution for signal levels at the output is represented by the lower PDF on the *Y* axis. PF slopes, which reflect the standard deviation of these underlying PDFs, would be steep for low-threshold conditions because the standard deviation of the distribution on the *X* axis would be small. In conditions that produce higher signal levels at masked threshold, the signal will pass through the more compressive region of the function, and will have to change by a greater amount (PDF on the right along the *X* axis) to produce the same change in signal level at the output of the nonlinearity (upper PDF along the *Y* axis). In this case, the PF slope would be more shallow (i.e., the underlying PDF would have a larger standard deviation). Internal noise is assumed to be uniform across conditions, and is reflected in this figure by similar standard deviations for the upper and lower PDFs on the *Y* axis.

dynamic nature of this relationship has tended to obscure the role of the peripheral nonlinearity and the model has provided valuable insights concerning the effects of specific changes in stimulus parameters or of changes in the nonlinearity itself (e.g., Oxenham and Plack, 1997). Psychometric functions (PFs), which describe the relationship between stimulus level and percent correct in a given condition, may provide a means of exploring the effect of the nonlinearity in a simpler framework in which only one quantity, either signal or masker level, is varied. In particular, the slope of the PF might be expected to provide a direct measure of the degree of nonlinearity at the signal place on the BM in the range of intensities traversed by the signal or masker.

For example, in variable-signal conditions, the combination of a moderate masker level and a longer signal delay produces a low signal level at masked threshold. In this case, the signal passes through the more linear portion of the BM I/O function as its level varies. Thus, changes in signal level at the input of the cochlear nonlinearity will result in similar changes in signal level at the output of the nonlinearity, and a smaller range of input signal levels will be presented in establishing threshold (Fig. 1). The underlying distribution of input signal levels in this case is represented by the probability density function (PDF) on the left along the *X* axis, and the corresponding distribution for signal levels at the output of the nonlinearity is represented by the lower PDF on the *Y* axis. PF slope is related to the PDF width such that the wider the PDF (i.e., the greater the standard deviation of the underlying distribution), the shallower the PF slope. Thus, associated PFs for low masked-threshold conditions would be steep because the standard deviation of the distribution on the *X* axis would be small. In conditions that produce higher signal levels at masked threshold, such as with shorter signal delay and/or higher masker level, the signal will pass through the more compressive region of the BM I/O function. The signal level at the input of the nonlinearity will have to change by a greater amount to produce the same change in signal level at the output of the nonlinearity, thus the PF slope would be more shallow. This relationship is illustrated by the wider PDF on the right along the *X* axis, and the corresponding upper PDF on the *Y* axis. Note that the range of the PF at the output of the nonlinearity is specified by internal noise, which is assumed to be uniform across conditions (Plack and Oxenham, 1998). This is illustrated by similar standard deviations for the upper and lower PDFs on the *Y* axis. We also would predict that PF slopes should be similar for on- and off-frequency maskers in the variable-signal case if the signal levels at threshold are similar for the two conditions. When signal level is varied in order to determine threshold at a fixed masker level, the effect of the nonlinearity should be independent of the properties of the masker.

In the current experiments 1 and 2, variable-signal conditions were presented that produced a range of signal levels at masked threshold for on- and off-frequency maskers. Mean thresholds from the on-frequency masker case in experiment 2 were used to estimate the parameters of the nonlinearity as described by Plack and Oxenham (1998) using data analysis software provided by Plack. The purpose was to demonstrate the ability of the Plack and Oxenham model to predict our own masked thresholds before extrapolating the model predictions to PFs. Next, PF slopes were estimated and compared across conditions. It was hypothesized that PF slope would decrease as a function of signal level, and that this trend would be similar for on- and off-frequency masker cases.

Although variable-masker conditions have been used frequently in forward masking studies (Nelson *et al.*, 2001, 1990; Nelson and Freyman, 1987, 1984; Oxenham and Plack, 1997), PFs have not been reported for those conditions. When the masker level is varied to determine threshold at a fixed signal level, the masker travels through the nonlinearity at the place of the signal, but the growth of the masker at the place of the signal is different for on- and off-frequency maskers. This is because the growth of the on-frequency masker will be influenced by the compressive nonlinearity at the place of the signal, whereas the growth of the off-frequency masker will be linear at the place of the signal. Thus, we predicted that PF slopes should be different for on- and off-frequency maskers in variable-masker conditions. Masker levels were varied to find threshold at several fixed signal levels for on- and off-frequency maskers in experiment 3 and PFs were fitted to those data.

It should be emphasized that the purpose of the current paper is to demonstrate that the peripheral nonlinearity has an effect on PF slopes, and not to identify a specific form of the underlying nonlinear function. More recent models that

incorporate a compressive nonlinearity have assumed a continuously level-dependent function, such as that employed in Glasberg and Moore (2000), rather than the two-straight-line approximation assumed in the Plack and Oxenham (1998) model. We focus on the latter model because its properties allow for specific predictions about forward-masked thresholds and associated PFs, and because we have used select conditions from Plack and Oxenham (1998) and Oxenham and Plack (2000) in our experiments 1 and 2.

All of the PFs reported here have been reconstructed from trial-by-trial data obtained with an adaptive procedure using data collection and analysis procedures described and tested by Dai (1995). These procedures provided an efficient method of obtaining PFs, while allowing us to obtain threshold estimates using a procedure comparable to that used by Plack and Oxenham (1998). As a result of using an adaptive procedure and then fitting a PF to the data, there are two ''threshold'' estimations described in the Results section for each experiment. The thresholds estimated from the adaptive procedures are presented first (i.e., the GOM functions), then the PF data are presented, which include slope and threshold from the PF fitting procedures. Where PF slopes are compared to threshold, ''threshold'' in this case refers to the values obtained from the PF fitting procedure, although the good agreement between these measures makes them interchangeable.

II. EXPERIMENT 1

Variable-signal conditions were presented for combinations of three on-frequency masker levels and three signal delays. PFs were reconstructed for each subject in each condition. If PF slope is determined by the degree of peripheral nonlinearity, and only by the degree of peripheral nonlinearity, then slopes should decrease as masked threshold for the signal increases.

A. Subjects

Three subjects, ages 20 to 41, participated in experiment 1 (S1–S3). Two of the subjects were the first and second authors. The other subject was a college student from a local university who was paid for his participation. This subject was informed of the procedures and purpose of the study, and signed a consent form prior to participation. Hearing for the authors was screened at 0.5, 1.0, 2.0, and 4.0 kHz in both ears, using laboratory stimulus generation and threshold measures as described in the following sections. Thresholds were at or better than 20 dB SPL at each test frequency, bilaterally. Hearing for the third subject was screened using standard clinical procedures as part of a research project for another laboratory at this facility. Results from that test indicated thresholds were at or better than 15 dB HL at octave and interoctave test frequencies from 0.25 to 8.0 kHz, bilaterally.

B. Stimuli

The variable-signal conditions in the current experiment 1 were a subset of the conditions used by Plack and Oxenham (1998) in their experiment 1. The signal was a 2-kHz, 10-ms tone (5-ms rise/fall) shaped with a cosine-squared envelope. The forward masker was a 2-kHz, 20-ms tone (2-ms rise/fall) shaped with a cosine-squared envelope, presented at 50, 70, or 90 dB SPL. Each masker level was presented in conditions in which the signal delay (measured at 0-voltage points) was 2, 20, or 40 ms. Threshold for the signal in quiet also was determined for each subject.

The masker and signal were generated digitally at 50 kHz (TDT AP2), and routed through separate 16-bit D/A converters (TDT DD1), then to an anti-aliasing filter (FT6) with a 20-kHz low-pass cutoff. The filtered stimuli were routed to programmable attenuators (PA4), mixed (SM3), then presented through headphone buffers (HB6) to the left ear of each participant through circumaural headphones (Sennheiser HD 250 II). Additional attenuation of up to 30 dB was available from resistor pads located in each booth. The headphones were plugged directly into boxes that contained these resistor pads.

C. Threshold determination

Thresholds for the signal in quiet and in the various experimental conditions were obtained using a 2IFC adaptive procedure, in which the level of the signal was decreased after two correct responses and increased after one incorrect response (Levitt, 1971). The step size was 8 dB until the fourth reversal, after which the step size was 4 dB. The step sizes were selected to reduce the number of levels visited by the adaptive procedure and increase the number of trials per level, an approach that Dai (1995) found to result in more stable estimates of PF parameters. Subjects responded using hand-held microterminals (Intelligent Instrumentation CTM280B). A trial consisted of a 500-ms warning interval (indicated by a ''*'' on the microterminal), two observation intervals (indicated by a ''1'' and ''2'') separated by 500 ms, an answer interval (indicated by a ''?''), and a 300-ms feedback interval in which the correct answer was displayed. There was a 500-ms interval between the trials, and 50 trials were presented in each block. Threshold for a block was calculated as the mean of the signal levels of all reversals after the fourth reversal. Conditions were presented in counterbalanced order. Subjects completed a minimum of 13 50-trial blocks for each of the nine conditions (three signal levels $\times$ three masker levels). Data were collected in 2-h sessions, and subjects were given breaks periodically during the sessions. Mean threshold for each subject in each condition was calculated using thresholds from the last 12 blocks of trials.

D. Psychometric functions

PFs have been analyzed in many different coordinates, particularly in the literature on intensity discrimination (e.g., Buus and Florentine, 1991; Moore *et al.*, 1999). In the literature on detection, the most common approach is to assume that the PF can be described by

$$d' = m(I)^k,$$

where the special cases of amplitude detection and energy detection are represented by $k=0.5$ and $k=1$, respectively

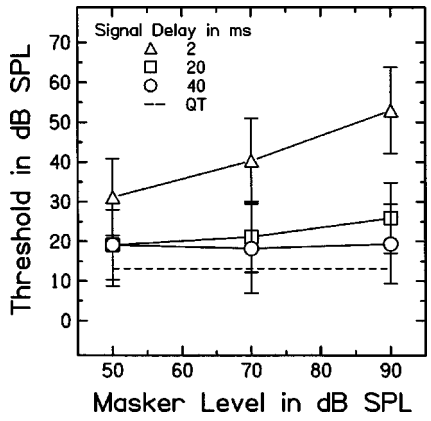


FIG. 2. Signal levels at masked threshold as a function of masker level, with signal delay as the parameter, for the mean across subjects for experiment 1. The forward masker was a 2-kHz, 20-ms tone presented at 50, 70, or 90 dB SPL and the signal was a 2-kHz, 10-ms tone presented at 2- (triangles), 20- (squares), or 40-ms (circles) signal delay in a variable-signal condition. Error bars represent plus and minus one standard deviation. The dashed line represents mean threshold for the signal in quiet (QT). Thresholds are low and slopes of GOM for the 20- and 40-ms delay conditions are shallow. Thresholds are higher in the 2-ms delay condition, and the slope of GOM is steeper for all subjects.

(Egan *et al.*, 1969). Other authors have assumed an underlying function of this form when simulating recovery of PF parameters from adaptive-tracking data (Dai, 1995; Leek *et al.*, 1992). Dai (1995) used a modified form of the equation

$$d' = (I/I_o)^k$$

where $I_o = (I/m)^k$ in the Egan *et al.* equation. In Dai's version, $10\log(I_o)$ represents the level required for $d' = 1$, whereas values of m in the original equation have no meaning independent of k. In Dai's approach, I_o and k are varied to minimize the weighted squared deviation between obtained and estimated values of percent correct, a χ^2 measure. Weights reflect the standard errors associated with the estimated percent correct values. The exponent k can be treated as a measure of the slope of the PF, given that $10\log d' = k*10\log(I/I_o)$. We have used Dai's version of the formula in the data analyses reported here.

The trial-by-trial data were pooled across the last 12 blocks to obtain estimates of percent correct as a function of signal level for each subject in each condition. Signal levels that were presented on 30 trials or less and/or produced less than 50% correct were not included in the fits.[1] The Dai fitting procedure provided the threshold and the slope of the PF for each condition for each subject. These parameters were averaged across the three subjects to obtain grand average PFs in each condition. The average across subjects included only those PF fits that had an r^2 of 0.50 or greater.

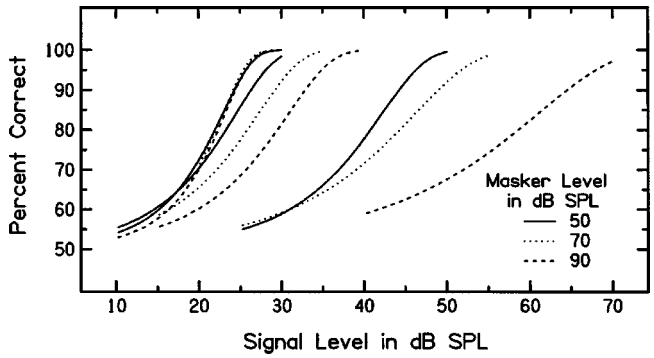


FIG. 3. Mean PFs for all nine conditions in experiment 1. The 50-dB masker conditions are represented by the solid lines, 70-dB maskers by dotted lines, and 90-dB maskers by dashed lines. Signal delay is not directly identified in this figure. As the signal levels increase, the slopes of the PFs decrease, indicating that presentation of a wider range of signal levels was required to determine threshold at higher levels.

E. Results

Mean signal levels at threshold across subjects as a function of masker level, with signal delay as the parameter, are displayed in Fig. 2. Error bars represent the standard deviation of thresholds across subjects. Masked thresholds for the 20- and 40-ms delay conditions are low, slopes of GOM (i.e., the rate at which masked threshold increases as a function of masker level) are shallow, and the two functions are similar. Masked thresholds are much higher in the 2-ms delay condition, and the slope of GOM is much steeper.

Although Dai's maximum-likelihood procedure for fitting PFs minimizes χ^2, unweighted r^2 values provide a more interpretable measure of goodness of fit. Table I provides a summary of the PF thresholds and slopes, and r^2 values for the individual PF fits. All but one fit met the criteria of $r^2 > 0.50$ for inclusion in the mean data (S1, 50-dB masker, 2-ms signal delay). Mean PFs for all nine conditions are shown in Fig. 3. Overall, as the signal levels increase, the slopes of the PFs decrease. Thus, presentation of a wider

TABLE I. PF thresholds and slopes, and r^2 values for PF fits for each subject in each condition in experiment 1.

Subject	Masker level	2-ms signal delay			20-ms signal delay			40-ms signal delay		
		Threshold	Slope	r^2	Threshold	Slope	r^2	Threshold	Slope	r^2
S1	50	17.98	0.10	0.40	12.24	0.65	0.97	9.63	0.84	0.95
	70	30.46	0.36	0.94	13.67	0.37	0.92	8.99	0.60	0.96
	90	45.96	0.28	0.76	19.16	0.39	0.89	9.16	0.65	0.88
S2	50	39.30	0.67	0.96	29.44	0.58	0.95	30.79	0.56	0.90
	70	48.06	0.60	0.97	32.27	0.54	0.88	31.33	1.11	0.99
	90	56.72	0.36	0.92	32.59	0.69	0.99	28.71	1.03	0.99
S3	50	39.45	0.39	0.93	24.38	0.58	0.93	22.45	0.92	0.99
	70	48.65	0.20	0.83	27.54	0.74	0.80	23.21	0.90	0.89
	90	65.84	0.28	0.82	32.54	0.54	0.94	26.50	0.94	0.96

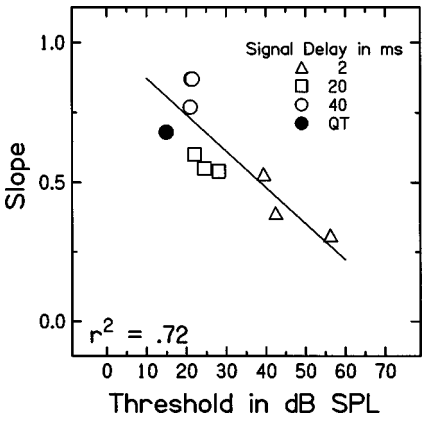


FIG. 4. Mean slopes of PFs as a function of mean threshold in dB SPL across subjects for all masker levels and signal delays in experiment 1. The filled circle represents the PF threshold and slope for the quiet threshold condition (QT). The r^2 value of 0.72 is from the linear regression fit. Note that this r^2 is associated with the regression line for slope versus threshold, not individual PF fits.

range of signal levels is required to determine threshold at higher levels. In Fig. 4, the mean slope of PFs across subjects is shown as a function of mean threshold for the nine masker level and signal delay combinations. PF slope decreases as masked threshold increases, and the relation is well described by a straight line. However, the data points for the two longer signal delays are not well separated, either in terms of slope or threshold. Shorter signal delays may have produced a better spread of masked thresholds and a more appropriate set of data for regression analysis.

Individual PF fits to the data for three of the nine conditions are shown in Fig. 5. As stated previously, the two longer signal delays produced mean masked thresholds that were very close together (see Fig. 2), and this is reflected in the PFs in the lower signal level range for individual subjects. The data points, particularly in the middle of the percent correct range, are well represented by the fits. However, values closer to 50% correct are weighted less in the fitting procedure, and sometimes are not well represented by the fitted line.

III. EXPERIMENT 2

Variable-signal conditions were presented using three shorter signal delays and a longer duration masker to produce more masking than in experiment 1. Off-frequency forward maskers also were included. The stimuli and signal delays were the same as those in experiment 1 of Oxenham and Plack (2000). Software supplied by Plack was used to estimate the nonlinearity parameters from the mean data from on-frequency masker conditions. It was predicted that PF slopes would decrease as a function of signal level at masked threshold, and that this trend would be similar for on- and off-frequency masker conditions.

A. Subjects

Three males and four females from 19 to 35 years of age participated in experiment 2 (S4–S10). Two of the males and one of the females were tested simultaneously in one group, and the remaining subjects were tested together in a second group. Each participant was informed of the procedures and purpose of the study, and signed a consent form prior to participation. All participants were college students from a local university, and none had previous experience with the task. They were paid for their participation. Hearing was screened at 0.5, 2.0, and 4.0 kHz in both ears using the same procedures as described for the authors in experiment 1. Thresholds for all participants were at or better than 20 dB SPL at each test frequency, bilaterally.

B. Stimuli

The signal was a 4-kHz, 10-ms tone (5-ms rise/fall) shaped with a cosine-squared envelope, and presented at delays of 0, 10, or 30 ms. The on-frequency forward masker was a 200-ms tone, presented at 30 to 90 dB SPL in 10-dB steps in different blocks. The off-frequency forward masker was a 2.4-kHz, 200-ms tone presented at 60 to 90 dB SPL in 10-dB steps in different blocks. Both maskers were shaped with 2-ms cosine-squared rise/fall envelopes. Stimulus generation and delivery was the same as for experiment 1, with

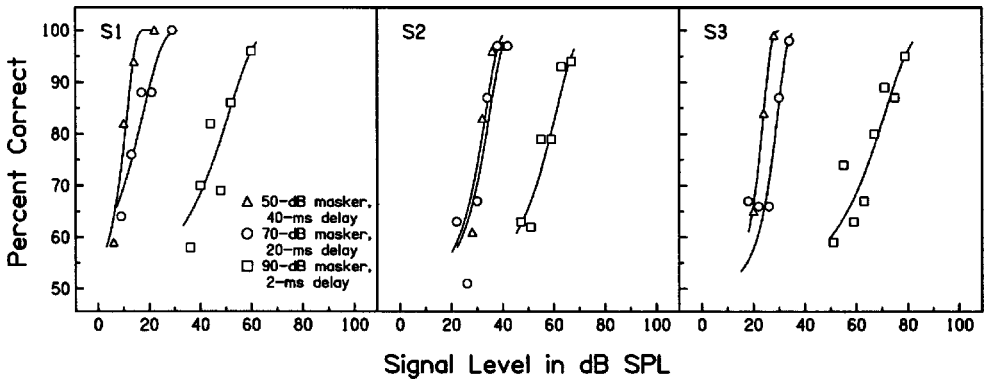


FIG. 5. Individual PF fits for select variable-signal conditions for experiment 1. The open symbols represent the actual data points, and the solid lines represent the fits to the points. The parameter and r^2 values are listed in Table I. Figure 2 shows that the two longer signal delays produced mean masked thresholds that were very close together, and this is reflected in the PFs for individual subjects. The fits to the data points are good, particularly in the middle of the percent correct range. Values at the lower end of the percent correct scale tend to be weighted less in the fitting procedure, and sometimes are not well represented by the fitted line. In these individual fits, particularly for S3, it can be seen that PF slopes tend to decrease as signal level increases.

the exception that multiple sets of attenuators, mixers, and headphone buffers were used to enable simultaneous testing of up to four subjects.

C. Threshold determination

Thresholds were determined using the same procedures as experiment 1, with the exception that step sizes were all 4 dB. The first group (S5, S7, and S9) completed 12 50-trial blocks of each condition; the second group (S4, S6, S8, S10) completed eight blocks per condition. Data from all subjects were combined and treated as one group for subsequent analyses.

D. Data analysis

Plack and Oxenham (1998) used temporal window parameters established by Oxenham and Moore (1994) and nonlinearity parameters established by Oxenham and Plack (1997) to fit their data. After adjusting the breakpoint of the nonlinearity from 45 dB (used by Oxenham and Plack, 1997) to 35 dB, they varied only the efficiency parameter, *k*, the threshold-level difference at the output of the temporal window required for detection of the signal (Plack and Oxenham, 1998). The model assumes that *k* is constant across conditions. In a later study, Oxenham (2001) varied the lower (more linear) and upper (more compressive) slope of the nonlinearity as well as window parameters to obtain an optimum fit, but allowed only a few parameters to vary at once. In fitting the current data, we have taken a similar approach. Using the standard window parameters, two successive iterations were performed to arrive at best estimates of the nonlinearity parameters based on the group mean adaptive thresholds in the on-frequency conditions.[2] First, lower slope was fixed at 0.78 and *k*, breakpoint, and upper slope were allowed to vary until a least squares fit was obtained. The resultant values for *k*, breakpoint, and upper slope were then fixed, and lower slope was allowed to vary until a least squares fit was obtained.

PFs were calculated for each subject in each condition using the Dai (1995) procedure. Signal levels with 30 trials or less and/or less than 50% correct were excluded from the fits. Of the 231 PFs (excluding those for quiet threshold) for seven listeners and 33 conditions, 218 were used in further analyses. Twelve PFs were excluded because of poor fits ($r^2 < 0.50$), and one was excluded because only one signal level remained after the trim procedure. Poor fits typically resulted from nonmonotonic functions. Correlations were calculated using SPSS to examine the relationship between mean PF slope and mean threshold from the PF fits, masker level, signal delay, and masker frequency.

E. Results

Mean signal levels at masked threshold (across replications, then across subjects) are shown as a function of masker level with signal delay as the parameter in Fig. 6. In the on-frequency masker conditions, the shorter signal delays and longer masker duration used in experiment 2 produced higher signal levels at masked threshold in comparison to experiment 1. In both the on- and off-frequency conditions,

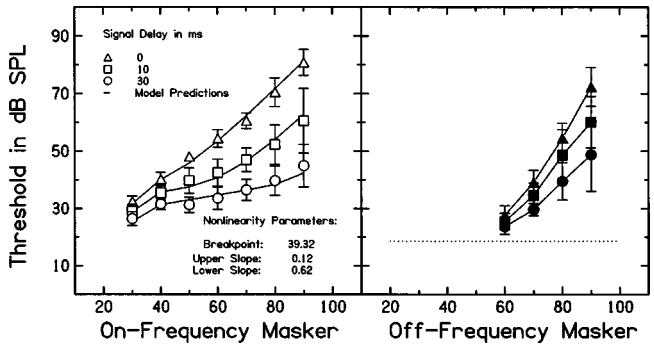


FIG. 6. Mean signal levels at threshold as a function of masker level with signal delay as the parameter for on-frequency (left panel) and off-frequency (right panel) masker conditions for experiment 2. The on-frequency forward masker was a 4-kHz, 200-ms tone presented at 30 to 90 dB SPL in 10-dB steps, and the off-frequency forward masker was a 2.4-kHz tone presented at 60 to 90 dB SPL in 10-dB steps. The signal was a 4-kHz, 10-ms tone presented at 0- (triangles), 10- (squares), or 30-ms (circles) signal delay in a variable-signal condition. Error bars represent plus and minus one standard deviation across subjects. The solid lines in the left panel for the on-frequency masker conditions represent the model fit to the mean data obtained with the specified parameters. The dotted line in the right panel represents threshold for the signal in quiet.

signal level at masked threshold increases with increases in masker level. However, the slopes of GOM are much steeper for the off-frequency maskers, and remain steep as the delay increases, whereas the slopes of GOM tend to decrease with an increase in signal delay for the on-frequency masker case.

The final fits of the Plack and Oxenham model to the mean on-frequency masking data also are shown in the left panel of Fig. 6. The rms error was 1.30 dB. The lower slope that produced the best fit was 0.62 dB/dB. This value is lower than the 0.78 dB/dB slope that Plack and Oxenham use, and physiological data suggest a value closer to 1.0. The upper slope of 0.12 dB/dB and breakpoint of 39 dB agree more closely with the values of 0.16 dB/dB and 35 dB assumed by Plack and Oxenham. The efficiency factor, *k*, was 2.58 dB.

The PF thresholds and slopes estimated from the Dai procedure and the corresponding r^2 values for a subset of the conditions are summarized in Table II. Because there is a large amount of data, only the on-frequency, 30-, 60-, and 90-dB masker conditions, and the off-frequency 60- and 90-dB masker conditions for each signal delay are presented to demonstrate the range of parameters and goodness of fit to individual subject data. PF fits to the individual data for a sample of low and high masked threshold conditions are presented in Fig. 7. PF fits whose parameters are included in Fig. 8 and in other summaries are shown by solid lines, and PF fits that are excluded are shown by dotted lines. These excluded conditions are the on-frequency, 60-dB masker, 30-ms signal delay condition for subject S7, and the on-frequency, 90-dB masker, 0-ms signal delay condition for subjects S6, S8, and S10.

Figure 8 shows the PF slopes plotted as a function of the PF thresholds for individual subjects, and the mean PF slope as a function of mean PF threshold across subjects. Note that subject S8 has extreme slope values (i.e., greater than 3.00) in six conditions, and they are not plotted here in order to keep the *Y*-axis scale in a range that is appropriate for the

TABLE II. PF thresholds and slopes, and r^2 values for PF fits for each subject for select conditions in experiment 2.

Subject	Masker frequency	Masker level	0-ms signal delay			10-ms signal delay			30-ms signal delay		
			Threshold	Slope	r^2	Threshold	Slope	r^2	Threshold	Slope	r^2
S4	On	30	34.60	0.97	0.86	29.84	0.82	0.90	24.76	1.41	0.99
		60	54.33	0.88	0.97	41.08	0.65	0.99	30.22	0.75	0.93
		90	82.76	0.24	0.64	63.28	0.24	0.80	49.76	0.38	0.46
	Off	60	25.12	0.91	0.99	23.19	1.10	0.85	20.85	1.33	0.98
		90	81.79	0.48	0.82	63.10	0.35	0.92	43.37	0.29	0.98
S5	On	30	34.68	0.91	0.75	32.92	1.30	1.00	30.67	0.49	0.54
		60	56.14	1.19	0.97	47.62	0.92	0.99	36.41	1.13	1.00
		90	85.80	0.47	0.83	58.02	0.22	0.80	44.82	0.96	0.96
	Off	60	27.50	1.06	0.98	29.08	1.45	1.00	28.99	1.17	0.98
		90	69.06	0.38	0.93	57.45	0.33	0.81	44.39	0.61	0.83
S6	On	30	34.68	0.98	0.93	31.34	0.98	0.92	30.20	0.76	0.91
		60	54.22	0.60	0.89	42.00	0.33	0.97	34.98	1.03	0.95
		90	76.30	0.10	0.37	50.42	0.64	0.97	41.45	0.64	0.87
	Off	60	29.65	1.47	0.99	26.96	2.12	0.99	24.57	0.99	0.92
		90	64.84	0.53	0.95	52.31	1.04	0.99	One-point	function	
S7	On	30	21.41	0.16	0.21	28.05	2.23	0.99	29.32	0.28	0.58
		60	57.19	0.35	0.53	36.83	0.79	0.91	30.05	0.08	−0.13
		90	88.17	0.30	0.96	49.23	0.13	0.61	36.55	0.70	0.89
	Off	60	34.16	0.29	0.52	29.91	0.35	0.85	26.15	0.72	0.94
		90	77.57	0.26	0.62	54.54	0.43	0.83	45.03	0.30	0.70
S8	On	30	31.20	2.23	0.98	29.56	4.19	1.00	27.15	4.20	0.98
		60	58.72	2.02	0.97	44.66	0.75	0.86	34.06	1.72	1.00
		90	86.33	0.24	0.41	66.60	0.25	0.95	44.52	0.23	0.74
	Off	60	29.57	0.66	0.86	29.57	1.29	1.00	27.59	1.19	0.95
		90	78.63	0.94	0.96	71.34	0.27	0.66	54.12	0.45	0.68
S9	On	30	27.80	0.26	0.99	33.77	0.88	0.89	31.27	0.61	0.87
		60	61.39	0.66	0.85	50.07	0.57	0.96	39.27	0.51	0.92
		90	83.73	0.25	0.68	69.65	0.19	0.84	55.83	0.27	0.88
	Off	60	28.82	1.07	0.95	28.04	0.66	0.88	27.23	1.02	0.69
		90	76.65	0.38	0.88	67.67	0.30	0.89	51.82	0.26	0.69
S10	On	30	37.53	0.62	0.98	33.73	0.54	0.93	29.06	0.33	0.94
		60	58.61	0.45	0.99	52.53	0.40	0.88	42.27	0.28	0.91
		90	85.52	0.06	0.18	87.78	0.24	0.66	60.08	0.42	0.83
	Off	60	34.66	0.35	0.79	25.33	0.48	0.98	25.35	0.53	0.91
		90	81.94	0.63	0.92	80.14	0.33	0.83	81.11	0.56	0.98

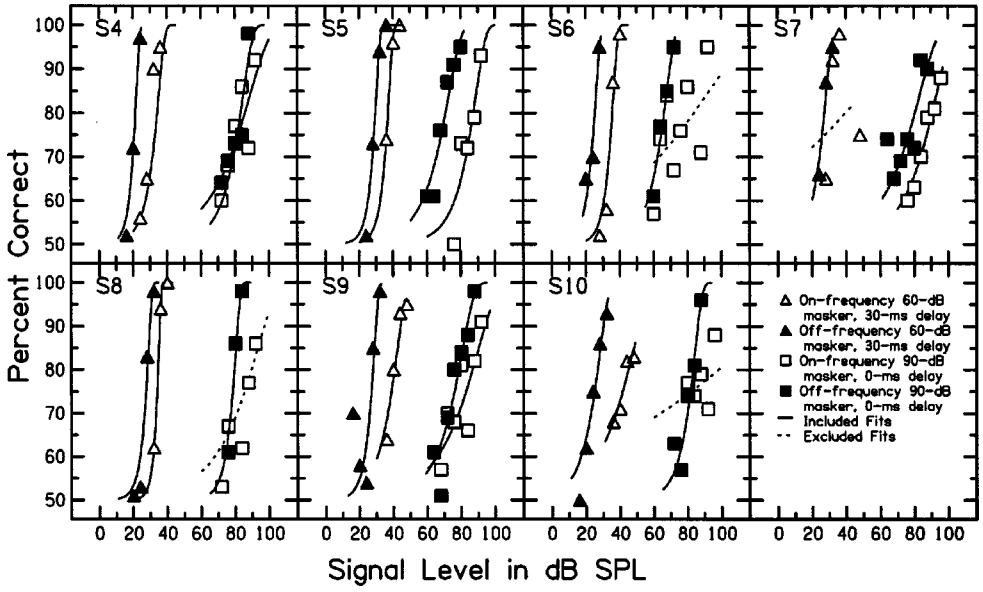


FIG. 7. PF fits to the individual data for a sample of low threshold (60-dB masker, 30-ms signal delay, represented by triangles) and high threshold conditions (90-dB masker, 0-ms signal delay, represented by squares) for on-frequency (open symbols) and off-frequency masker cases (filled symbols) for experiment 2. PF fits that were excluded from summaries due to $r^2<0.50$ are shown by dotted lines.

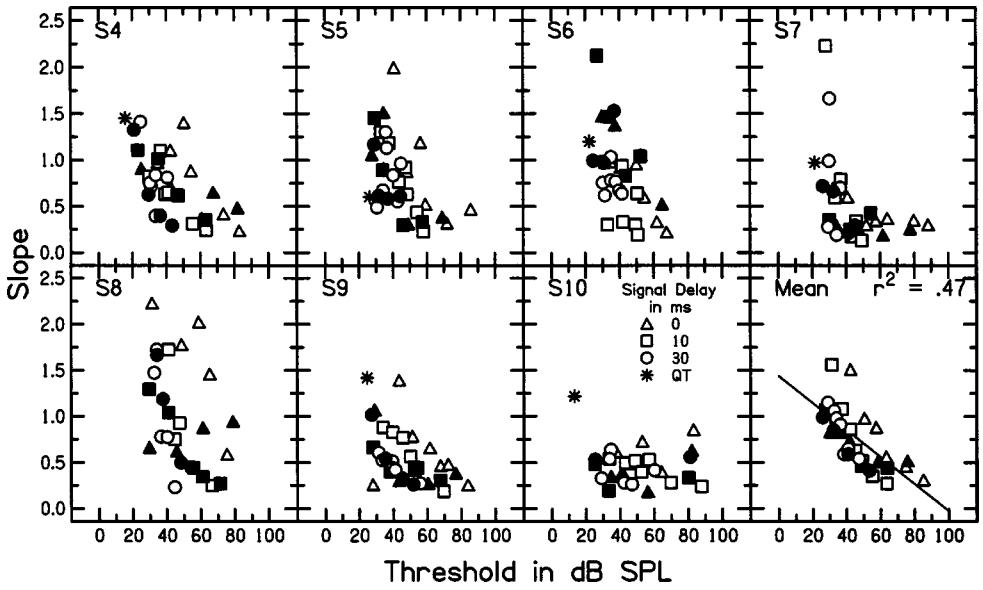


FIG. 8. Slopes of PFs plotted as a function of threshold for individual subjects, and the mean PF slope as a function of mean threshold for experiment 2. On- and off-frequency masker conditions are represented by open and closed symbols, respectively. The 0-ms signal delay conditions are represented by triangles, 10-ms delay by squares, and 30-ms delay by circles. Masker level is not directly identified in this plot. Threshold for the signal in quiet is represented by the star. Slopes of PFs decrease as signal threshold increases, and there does not appear to be a systematic difference between masker frequency conditions. However, there does appear to be a large amount of variability across subjects. Subject S10 has relatively uniform PF slopes across conditions. Also, subjects S4, S5, and S8 have higher signal thresholds in the on-frequency, 0-ms condition (open triangles) without a corresponding decrease in the slopes of the PFs. It should be noted that subject S8 has extreme slope values (i.e., >3.00) for five of the low-level, on-frequency masker conditions, and for the threshold in quiet condition. These data points are not plotted because the *Y* axis would have to be expanded to accommodate them, making the trends in the data for other subjects difficult to distinguish.

remaining subjects. For this subject, the slope values in the 40-dB, on-frequency masker conditions are 4.13, 3.06, and 3.55 in the 0-, 10-, and 30-ms signal delay conditions, respectively. Slope values in the 30-dB, on-frequency masker condition are 4.19 and 4.20 for the 10- and 30-ms signal delay conditions, respectively, and the slope value for threshold for the signal in quiet is 3.02. All of these conditions have associated PF thresholds that are below 41 dB SPL. In general, PF slopes decrease as signal level at threshold increases as observed in experiment 1. Further, there does not appear to be a systematic difference between on- and off-frequency masker conditions. However, there does appear to be a large amount of variability across subjects. Subject S10, for example, has relatively uniform PF slopes across conditions. Also, although the mean data do not indicate a systematic difference among signal delays and masker frequencies, subjects S4, S5, and S8 have higher masked thresholds in the on-frequency, 0-ms condition (open triangles) without a corresponding decrease in the slopes of the PFs.

The correlation of mean PF slope with mean PF threshold in Fig. 8 is -0.69. We also evaluated the correlation of mean PF slope with the stimulus parameters. The correlation is -0.89 with masker level, 0.06 with signal delay, and 0.18 with masker frequency. Only the correlations between PF slope and PF threshold and between PF slope and masker level are significant ($p<0.05$). PF threshold also is significantly correlated with masker level (0.66) and with signal delay (-0.49). The higher correlation of PF slope with masker level than with signal level is unexpected, and may be due in part to results obtained for the 0-ms signal delay, on-frequency masker condition. When subjects S4, S5, and S8 are excluded from the mean data for the 0-ms signal delay, on-frequency masker condition, the correlation of mean PF slope is -0.77 with mean PF threshold, -0.78 with masker level, 0.28 with signal delay, and 0.49 with masker frequency.

Given the form of the nonlinearity assumed by Plack and Oxenham (1998), described by two straight lines in dB coordinates, all PFs with signal levels falling only on the lower slope should be uniformly steep, while all PFs with signal levels falling only on the upper slope should be uniformly shallow. The pattern of results in Fig. 8 suggests a smooth nonlinearity that is continuously decreasing in slope and is consistent with recent models that incorporate smooth nonlinear functions (e.g., Glasberg and Moore, 2000) and with data that provide independent support for those models (e.g., Nelson *et al.*, 2001).

IV. EXPERIMENT 3

Variable-masker conditions were presented for on- and off-frequency maskers at one signal delay. In this case, it was predicted that the slopes of PFs for the on-frequency maskers would be shallow at higher masker levels, and that slopes of PFs for the off-frequency maskers would be steep across masker levels.

A. Subjects and procedures

Three females and one male from experiment 2 participated in experiment 3 (S4, S6, S8, and S10). The signal was a 4-kHz, 10-ms tone (5-ms rise/fall) shaped with a cosine-squared envelope, presented 0 ms after the offset of the masker. The on-frequency forward masker was a 4-kHz, 200-ms tone. The off-frequency forward masker was at 2.4

kHz. Maskers were shaped with 2-ms rise/fall cosine-squared envelopes. Stimulus generation and delivery were the same as for experiments 1 and 2.

Initially, signal levels were fixed at 30 and 50 dB SPL for both on- and off-frequency maskers. Thresholds were obtained using a 2IFC adaptive procedure in which the level of the masker was decreased after two correct responses and increased after one incorrect response. Step sizes were 4 dB. Threshold for a block was calculated as the mean of the masker levels of all reversals after the fourth reversal. Sixteen blocks of each condition were presented in a counterbalanced order. There was an extended break after completing data collection for these four conditions. In the process of data analysis, an error in the stimulus file was discovered for the off-frequency masker, 30-dB SPL signal level condition. All blocks in this condition were repeated for all subjects, along with four more blocks of the other three conditions as a check for consistent thresholds. The first four blocks that were collected previously in these conditions were excluded from further analysis.

The initial data for masker levels at threshold for fixed signal-level conditions of 30 and 50 dB were used to select signal levels that would produce an overlap in the range of masker levels at threshold for on- and off-frequency conditions. The final data set included threshold for the signal in quiet, and masked thresholds for fixed signal levels of 30, 50, 57, and 63 dB SPL for the on-frequency masker case, and 25, 30, 40, and 50 dB SPL for the off-frequency masker case. Twelve 50-trial blocks were completed for threshold in quiet, and 16 50-trial blocks were completed in the remaining conditions, with the exception of the off-frequency masker in the 25-dB SPL signal level condition. S6's threshold for the signal in quiet was 24.9 dB SPL, and she reported difficulty hearing the signal and masker in the 25-dB SPL fixed-signal condition. The condition was terminated after eight blocks for all subjects.

Subjects occasionally produced masked thresholds that clearly deviated (e.g., by more than 8 dB) from the rest of their masked thresholds in a particular condition. These data were excluded, and runs for the corresponding conditions were repeated. In one instance, a subject produced unusually variable thresholds for the 63-dB SPL fixed-signal, on-frequency masker condition (8.95 dB standard deviation across the 16 thresholds), which were all collected during one test session. These initial thresholds were discarded, and 16 more repetitions were collected. This second set of thresholds had a lower standard deviation, and were used for further data analysis.

PFs were calculated using the Dai (1995) method. However, because PFs in this case described percent correct as a function of masker level, percent correct increased as masker level decreased. PFs were fitted for individual subjects, and only those fits that produced r^2 values of >0.50 were included in further analyses. With this criterion, PFs were excluded for the off-frequency masker, 25 dB SPL signal condition for subjects S6 and S10. Correlations were calculated for mean PF slope as a function of mean PF threshold for on- and off-frequency masker conditions.

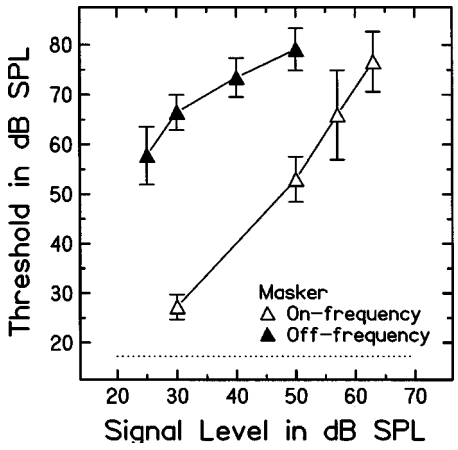


FIG. 9. Mean masker levels at threshold across subjects as a function of signal level for experiment 3. The on-frequency forward masker (open symbols) was a 4-kHz, 200-ms tone and the off-frequency forward masker (filled symbols) was a 2.4-kHz tone presented in a variable-masker condition. The signal was a 4-kHz, 10-ms tone presented at 0-ms signal delay. Mean threshold for the signal in quiet is represented by the dotted line. The signal was fixed in different conditions at 30, 50, 57, and 63 dB SPL for the on-frequency masker case, and at 25, 30, 40, and 50 dB SPL for the off-frequency masker case. Masker level at threshold is much higher in the off-frequency masker case than for the on-frequency case, particularly for the lower signal levels. The slope of GOM is more shallow for the off-frequency case. These trends agree with the literature.

B. Results

Mean growth of masking (GOM) functions in the variable-masker conditions are shown in Fig. 9. The GOM in the variable-masker conditions was steeper for on-frequency maskers than for off-frequency maskers. The PF thresholds, slopes, and r^2 values for each subject in each condition are summarized in Table III. The correlation of mean PF slope and mean PF masker level at masked threshold was -0.88 for the on-frequency masker case and -0.87 for the off-frequency masker case. Given the small number of data points, neither of the correlations was significantly different from zero. The negative correlation in the on-frequency case was expected, but the correlation was expected to be near 0 for the off-frequency case.

PF fits to the individual data for select conditions are shown in Fig. 10. As noted in experiments 1 and 2, there is variability in PF slopes among the subjects. For example, PFs for subjects S4 and S6 reflect the expected trends, that is, PFs are steeper in the off-frequency masker conditions across masker level in comparison to the shallower PFs in the on-frequency masker conditions. However, subject S8's PFs are of similar slope across all conditions. Subject S10's PFs are shallow across conditions, which is similar to experiment 2 in which the PFs for the same subject were shallow regardless of masker frequency, masker level, or signal delay (see Fig. 8). There are also concerns regarding individual fits. Although the PF fit for the off-frequency masker, 50-dB fixed-signal level condition for S10 met the criteria for inclusion in data analysis, for example, the line appears to approximate only one or two data points. Alternative fits with higher values of r^2 and better visual properties are available, but the values of χ^2 associated with those fits are much greater.

TABLE III. PF thresholds and slopes, and r^2 values for PF fits for each subject in each condition in experiment 3.

Subject	Masker frequency	Signal level	Threshold	Slope	r^2
S4	On	30	27.32	1.00	1.00
		50	54.53	0.55	0.99
		57	75.77	0.32	0.94
		63	79.33	0.51	0.89
	Off	25	61.71	1.75	1.00
		30	65.93	1.15	0.99
		40	72.79	1.51	1.00
		50	77.78	0.99	0.96
S6	On	30	26.42	0.62	0.97
		50	51.42	0.33	0.85
		57	60.96	0.26	0.88
		63	77.33	0.26	0.82
	Off	25	32.88	0.11	0.28
		30	64.39	1.38	0.98
		40	72.70	0.51	0.71
		50	80.41	1.13	0.99
S8	On	30	26.89	0.74	0.90
		50	50.97	1.18	0.90
		57	58.44	0.46	0.91
		63	67.86	0.45	0.81
	Off	25	53.33	0.55	0.82
		30	59.60	0.64	0.89
		40	65.84	1.09	0.97
		50	71.88	0.92	0.99
S10	On	30	21.69	0.52	0.90
		50	43.65	0.40	0.84
		57	59.20	0.20	0.90
		63	68.83	0.41	0.99
	Off	25	58.45	0.08	0.27
		30	67.40	0.33	0.86
		40	73.97	0.19	0.77
		50	77.93	0.27	0.80

V. DISCUSSION

A. Methodological issues

Our confidence in the interpretation of the current PF data is limited by several issues that are ongoing topics of debate in the literature. The first is the use of adaptive-procedure data to estimate PF parameters. Simulation studies suggest that procedures with more than two intervals, combined with a decision rule that estimates a higher point on the PF, provide more efficient and more accurate estimates of the PF slope (e.g., Leek *et al.*, 1992; Leek, 2001). Dai (1995), using a similar approach based on extensive simulations, found that estimates of PF parameters were unbiased when obtained using the data collection and analysis procedures we have used, although he did not explore the alternative decision rules tested by Leek *et al.* (1992). While Dai (1995) evaluated the accuracy of both the PF slope and the PF threshold estimate, only the accuracy of the slope estimate is of any concern here. The PF threshold estimates are highly correlated with the estimates obtained by taking the mean of the reversal points in the adaptive procedure ($r=0.99$ for the 231 thresholds in experiment 2), a widely used threshold measure.

The second issue is the choice of equation used to fit the PFs and to estimate the slope parameter. The estimates of PF slope reflect the representation of the stimuli used in fitting the PFs, as well as the assumed form of the PF. Any of the approaches used in the detection literature would show shallower slopes at higher signal levels, but the outcome of a statistical analysis could be somewhat dependent on the measure of slope that was used. We converted signal levels to power and fitted the data using the same equation used to fit PFs for absolute thresholds (e.g., Watson *et al.*, 1972). This approach seemed most consistent with the literature, given that the data were obtained in a masking procedure where we had no valid measure of the effective level of the masker at the time the signal was presented. The Plack and Oxenham (1998) model assumes, however, that the task is equivalent to intensity discrimination where the decision variable is an internal representation of $\Delta I/I$ or $(I+\Delta I)/I$. Those two quantities produce identical threshold predictions when using the model, but they do not result in identical PFs. Use of $\Delta I/I$,

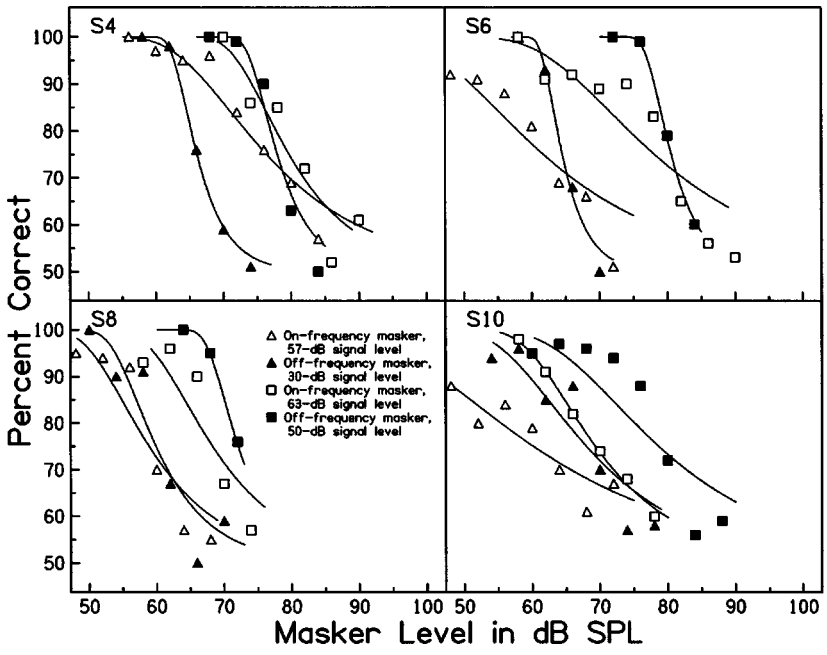


FIG. 10. Example PF fits to the individual data for variable-masker conditions in experiment 3. Data for the on-frequency masker, 57- and 63-dB fixed-signal-level conditions are represented by open triangles and squares, respectively. Data for the off-frequency masker, 30- and 50-dB fixed-signal-level conditions are represented by filled triangles and squares, respectively. PF fits to the data are represented by solid lines. PFs for subjects S4 and S6 demonstrate steeper PFs in the off-frequency masker conditions across masker level, and shallower PFs in the on-frequency masker conditions. However, PF slopes are similar across conditions for subjects S8 and S10s. For S10, PFs are shallow across conditions for variable-masker conditions (this figure) and variable-signal conditions (see Fig. 8).

where I was a constant dependent on masker level, would rescale all points on the function leaving the slope unchanged and thus would produce slope estimates identical to those we have obtained. Conversion to $(I+\Delta I)/I$ would be a nonlinear transform, with unclear results, but it seems unlikely that the differences in the fits would be large. We assumed a form of the PF (e.g., Egan *et al.*, 1969) that has been widely used in the detection literature, but many alternatives have been proposed. There are only two other studies in the literature that we are aware of that have reported PFs for forward masking conditions. Werner (1999) plotted PFs in percent yes votes versus signal level in dB, but did not fit an equation to the data. Nizami and Schneider (1999) used probit analysis to fit PFs as normal ogives in percent correct versus signal level in dB. Their PFs were similar in form to those generated with the Egan *et al.* equation.

The quality of some of the individual PF fits also is an issue. In order to avoid ambiguity in the analyses and the potential for selective data trimming when fitting a large number of PFs, we adopted strict rules for removing data points before fits were done and for excluding unreliable fits when computing mean slope values. The criterion of $r^2 > 0.50$ used for inclusion of the PF slope in summary data was somewhat arbitrary, but use of other criterion points did not have a strong effect on the correlation of slope with signal level or masker level. Some fits appeared to be less than optimal when the PFs were plotted with the actual data points (e.g., Fig. 10, S10, off-frequency masker, 50-dB signal level data) because the maximum likelihood procedure (Dai, 1995) assigned more weight to points higher on the PF.

Finally, it is a concern that the individual data were highly variable across subjects and conditions. For example, subject S10 demonstrated shallow PFs as a function of signal level, regardless of masker frequency, masker level, signal delay, or varied parameter (see Fig. 8 and Fig. 10). We do not have an explanation for this trend. One possibility is that this subject has greater internal noise in comparison to the other subjects. If this is the case, the difference in internal noise did not extend to detection of the signal in quiet, where results are in line with the remaining subjects (see the data points represented by stars in Fig. 8). Another example of individual variability is that a few of the subjects demonstrated higher PF slopes than would have been expected, given the associated signal levels, in the on-frequency masker, 0-ms signal delay, variable-signal condition (see S4, S5, and S8 in Fig. 8). There is not sufficient information to determine whether the slopes or the thresholds were unusually high. It also cannot be determined at this point whether this threshold/slope effect is one that should be taken into account in a model of forward masking, or could be attributed to a phenomenon such as confusion (Neff, 1985, 1986) that is outside the framework of the model. A confusion effect seems unlikely given that such effects are not generally found for sinusoidal maskers and signals, and tend to be associated with nonmonotonic PFs, with sharp decreases in percent correct in regions where the signal and masker levels are nearly equal. Of the individual PFs that met the criteria for inclusion in the mean, only one function (subject S4, 90-dB masker condition, see Fig. 7) displayed this trend.

B. Effects of the peripheral nonlinearity

Despite the methodological considerations discussed in the previous section, the peripheral nonlinearity is clearly reflected in both the GOM and PF data. In general, GOM functions for on- and off-frequency maskers and with various signal delays were consistent with previous reports in the literature for variable-signal conditions (Fastl, 1979; Jesteadt *et al.*, 1982; Kidd and Feth, 1981; Luscher and Zwislocki, 1947; Munson and Gardner, 1950; Oxenham and Plack, 2000; Plack and Oxenham, 1998; Widin and Viemeister, 1979; Zwislocki *et al.*, 1959) and variable-masker conditions (Nelson *et al.*, 2001, 1990; Nelson and Freyman, 1987, 1984; Oxenham and Plack, 1997). The Plack and Oxenham (1998) model, which includes peripheral nonlinearity as a first stage, provides an excellent account of the present GOM data for on-frequency maskers (see the left panel of Fig. 6). The current data for off-frequency maskers show a larger effect of signal delay on GOM functions than observed by Oxenham and Plack (2000), but the effect was still smaller than that observed with on-frequency maskers, in agreement with their model.

Plack and Oxenham (1998) make no specific predictions concerning PFs, but the assumptions of the model suggest that PF slopes should reflect the nonlinearity, and the model and supporting data specify the degree of nonlinearity expected for the on- and off-frequency maskers used here. The model assumes that internal noise is Gaussian in power with a standard deviation that is proportional to the mean (i.e., constant in dB), except in conditions where the masked threshold is near quiet threshold. Given that internal noise is constant, the slope or exponent of the PF should be proportional to the degree of nonlinearity. To a first approximation, the results of the present experiments were consistent with this prediction. In the variable-signal conditions of experiments 1 and 2, PF slopes decreased as a function of signal level for on- and off-frequency maskers. This was expected because in variable-signal conditions, the signal passes through a range of input levels to the nonlinearity at the place of the signal during threshold estimation. Thus, the associated PF slopes should be influenced by the nonlinearity at the place of the signal, regardless of masker frequency. Given this trend, it also was expected in experiment 2 that PF slope would be more highly correlated with signal level at threshold than with the other variables (i.e., masker level, signal delay, and masker frequency). However, the correlation of PF slope with masker level was greater than the correlation with signal level. This relationship is explored further in the following section.

In the variable-masker conditions of experiment 3, it was expected that PF slope would be shallower at higher masker levels for the on-frequency masker case, and that PF slope would remain steep across all masker levels for the off-frequency masker case. This is because in the variable-masker condition, masker level passes through a range of input levels to the nonlinearity at the place of the signal as the PF is obtained for any given condition. Thus, the on-frequency masker will be influenced by the nonlinearity at the place of the signal, and the associated PF slopes will reflect the amount of compression, whereas the off-frequency

masker will grow linearly at the place of the signal, and the associated PF slopes should be similar over the entire range of masker levels. Only two of the four subjects demonstrated this trend. Because on- and off-frequency masker cases were examined separately, there were fewer data points available for the correlation in experiment 3 in comparison to experiment 2 in which on- and off-frequency data were combined. Given the small number of data points, we cannot draw any meaningful conclusions from the correlation coefficients for experiment 3.

Both of the previous studies that describe PFs for forward-masked conditions used variable-signal procedures. Werner (1999) used a 65-dB SPL broadband noise to mask a 1-kHz tone with signal delays varying from 5 to 200 ms in normal-hearing adults and infants 3 to 6 months of age. The PFs for the adult group were steep and parallel across signal delay, and all PFs were restricted to signal levels below approximately 40 dB SPL. Thus, the PFs for the adult group in the Werner study were not in a signal-level range that would have reflected the effects of the nonlinearity as in the current study. Even if the PFs had extended into a higher signal range, it is not clear that the effects of the nonlinearity should be evident in conditions using a broadband masker rather than a pure tone masker. Nizami and Schneider (1999) obtained their data using a 97-dB, on-frequency, 2-kHz tone to mask a Gaussian-shaped signal, and manipulating signal delay in very small steps. While we found a large effect of masker level on PF slope in experiment 2, in the Nizami and Schneider study, masker level was constant. Their data point to a strong effect of signal level. They also observed fine structure and nonmonotonicity in their PF parameters, characterized by sharp increases in masked threshold and slope with successive increases in signal delay. This fine structure was not observable in the current data because of the large differences in signal delays (2, 20, and 40 ms in experiment 1, and 0, 10, and 30 ms in experiment 2 in the current study, as opposed to 0.5-ms steps in some cases in the Nizami and Schneider study). The general trend in their data was for masked threshold to decrease and PF slope to increase as signal delay increased. This is consistent with the current data (experiments 1 and 2, on-frequency masker) and with the interpretation that signal level at threshold determines the slope of the PF.

The Plack and Oxenham (1998) model can be extended to make specific quantitative predictions about PF slope ratios, in addition to the qualitative predictions reviewed above. If the slope or exponent of the PF is truly proportional to the degree of nonlinearity, then the compression ratio assumed in the model should be reflected in the slopes estimated from the data. Plack and Oxenham (1998) assumed a compressive nonlinearity with a lower slope of 0.78 dB/dB and an upper slope of 0.16 dB/dB, a ratio of 4.9 to 1 in the model fits to their on-frequency masker data. This is similar to the ratio of 5.2 to 1 that was obtained with the parameters estimated from the current GOM data (see Fig. 6, lower slope of 0.62 and upper slope of 0.12) and predicts a ratio of 5 to 1 in the slopes of the PFs. The observed ratio of the mean PF slope at 25.8 dB (slope=0.99), the lowest signal level in the mean data panel of Fig. 7, to the mean slope at 85.1 dB (slope=0.31), the highest signal level, was 3.2 to 1. The smaller ratio may reflect errors in fitting the PFs.

Finally, regardless of the strength of the relationship between signal level and PF slope, the pattern of PF slopes across conditions should predict the general form of the nonlinearity. The assumption in the Plack and Oxenham (1998) model that the nonlinearity can be approximated by a two-part function defined by two straight lines predicts two groups or clusters of data points in Figs. 4 and 8, one at high PF slope values (i.e., to reflect the linear portion of the function) and the other at low slope values (to reflect the compressive portion of the function). The PF slope data suggest a nonlinearity with a continuously decreasing slope or a very gradual transition, consistent with more recent models that include a peripheral nonlinearity (e.g., Glasberg and Moore, 2000).

The degree of peripheral nonlinearity has been estimated from masking data using the comparison of slopes of GOM for on- and off-frequency maskers in variable-masker conditions (Oxenham and Plack, 1997), from the ratio of temporal-masking-curve slopes for on- and off-frequency maskers (Nelson *et al.*, 2001), and from behavioral thresholds in on-frequency masking conditions with a range of signal delays and masker levels, such as those shown in Fig. 6. The current data suggest that measures based on slopes of PFs might offer another alternative that, with use of optimum procedures for measurement of PFs, could be both efficient and broadly applicable. The ratio of slopes of GOM functions for on- and off-frequency maskers only provides an estimate of the compressive portion of the nonlinearity, not of the transition or the lower, steeper portion of the function. In contrast, the ratio of fixed-probe temporal-masking-curve slopes for on- and off-frequency maskers provides an estimate of the transfer function for a larger range of input levels, including the lower levels (Nelson *et al.*, 2001). However, collection of a full set of data for a range of masker levels and signal delays is time consuming and recovery of nonlinearity parameters for either method requires a specific model with numerous assumptions. The range of PF slopes from a carefully selected set of conditions could be used to provide a more complete estimate of the nonlinearity, with fewer *a priori* assumptions concerning its form.

C. Exploring the masker level effect on PF slope

By far the biggest issue in the interpretation of the current data is the strong effect of masker level on the slope of the PF. The original hypothesis was that PF slope would have a higher correlation with signal level than with the other variables in experiment 2 because the signal level was varied in the adaptive procedure to reconstruct the individual PFs within conditions. However, it was found that PF slope was more highly correlated with masker level than with signal level. Interpreting the relationship between PF slope and signal level, and between PF slope and masker level, is complicated by the fact that masker level and signal level are highly correlated with each other. We examined these relationships further using partial correlations to obtain a measure of the relation between PF slope and each variable, with all other

variables held constant. The partial correlation of PF slope with threshold was −0.06, with masker level was −0.71, with signal delay was 0.05, and with masker frequency was 0.28. Thus, signal level at threshold had no almost relationship to PF slope once the remaining variables were accounted for in the comparison. Masker level had the strongest relationship with PF slope, indicating that as masker level increased, PF slope decreased with signal level held constant. This effect is not predicted by the model described in Fig. 1. When the data from the three subjects (S4, S5, and S8) in the 0-ms signal delay conditions were excluded from the analysis, the overall correlation of PF slope decreased with masker level (from −0.89 to −0.78) and increased with signal level (from −0.69 to −0.77). A similar trend was noted in the partial correlations of PF slope with masker level (a decrease from −0.71 to −0.61) and with signal level (an increase from −0.06 to −0.10). Partial correlations of PF slope increased with signal delay to 0.29 and with masker frequency to 0.34. This shift in the zeroth order and partial correlations cannot account for the masker level effect, and it does not change the conclusion that masker level has a strong, independent contribution to PF slope.

Three factors might contribute to the strong effect of masker level in experiment 2. First, the relationship between signal level at threshold and PF slope is not linear, and therefore would be poorly represented by a correlation that assumes a linear relationship between variables. The mean data panel in Fig. 7 could be viewed as nonlinear, particularly if the 0-ms signal delay conditions are disregarded, and the model described in Fig. 1 predicts a nonlinear effect. We do not have enough data points to support a nonlinear regression analysis. Second, the correlation of PF slope with signal level at threshold is attenuated by the measurement error associated with the latter variable. Masker level, masker frequency, and signal delay are all independent variables whose values are not subject to measurement error. We do not have an appropriate measure of the reliability of the estimates of signal level at threshold that could be used to correct for attenuation (e.g., Guilford, 1965, p. 487). Data from experiment 3, where masker level was the dependent variable, are too limited to provide a valid comparison.

Third, the strong effect of masker level might reflect an increase in internal noise with an increase in overall level. In the example provided in Fig. 1, the internal noise is assumed to be the same across conditions as represented by lower and upper PDFs with similar standard deviations on the *Y* axis. An increase in internal noise with increases in level, corresponding to the more compressive region of the function, would be represented by an upper PDF with a larger standard deviation. This would result in a stronger effect of masker level than of signal level because masker levels vary over a wider range than signal levels. We would also predict an effect of masker level with signal level held constant. Of the factors considered here, the assumption that internal noise is greater at higher levels seems best able to account for the observed pattern of results. This assumption would have a significant impact on procedures for fitting threshold data. It may be possible to test the assumption by measuring the amount of internal noise (e.g., Spiegel and Green, 1981).

VI. CONCLUSIONS

Growth of masking results for on- and off-frequency forward-masker cases, in variable-masker and variable-signal conditions, were consistent with the literature. The mean on-frequency masker data from experiment 2 were well represented by a model described by Plack and Oxenham (1998) that incorporates a peripheral nonlinearity as its first stage. The model assumptions suggest that this nonlinearity also should be reflected in the PF slopes. Mean PF slopes decreased with increases in mean signal level at threshold for on- and off-frequency masker cases in the variable-signal conditions of experiments 1 and 2. PF slopes were shallow for on-frequency maskers and steep for off-frequency maskers at high thresholds for two of the four subjects in the variable-masker conditions of experiment 3. These results are consistent with the hypothesis that the peripheral nonlinearity is reflected in PF slopes, particularly for experiments 1 and 2. Further data collection would be necessary to make stronger statements about the influence of the peripheral nonlinearity on PF slopes for variable-masker conditions. PF slopes from select conditions could be used to produce more direct estimates of the specific form of the nonlinearity than can be obtained by fitting current models to threshold data. An unexpected result was the strong relationship of masker level with PF slope, independent of signal level. This finding might be explained by assuming that internal noise increases with increasing stimulus level. Greater internal noise would result in shallower PFs. Given that masker levels are higher than the corresponding signal levels at threshold in forward masking and that the range of masker levels is greater than the range of signal levels, an internal noise that increased with stimulus level would result in a correlation between PF slope and masker level that was higher than the correlation between PF slope and signal level. Further work will be required to determine the cause of the masker level effect.

ACKNOWLEDGMENTS

This work was supported by R01 DC00136 and by T32 DC00013. We thank Andrew Oxenham and Christopher Plack for extensive input regarding the model and Christopher Plack for providing a copy of the MatLab code used to fit the data shown in Fig. 6. We also thank Robert Lutfi for his instructive comments regarding the role of internal noise in forward masking and Huanping Dai for advice on obtaining and fitting psychometric-function data.

[1]The number of signal levels excluded during the PF fitting procedure varied among subjects, and depended, for example, upon the starting level of the signal at the beginning of each adaptive track. Because we collected data from three to four subjects simultaneously in experiments 2 and 3, the starting level was the same for all subjects. A greater number of signal levels were trimmed due to rare occurrence for subjects who had lower thresholds because they produced correct responses at a larger number of the initial signal levels in comparison to subjects with higher thresholds. As threshold was approached, many of these initial signal levels were not presented again for those with lower thresholds, but were presented repeatedly for those with higher thresholds. We counted the number of signal levels that remained after trimming the levels that were associated with low occurrence or less than 50% correct, and averaged these value across conditions and subjects for experiments 1 and 2. The mean number of signal

levels left after the trim was 5.1 (sd=1.3) for experiment 1, and 4.5 (sd =1.5) for experiment 2.

[2]The temporal window parameters used in fitting the current data were fixed at Tb1=4.00, Tb2=29.00, w=0.16, and Ta=3.50. Tb1, Tb2, and Ta were the same values Plack and Oxenham (1998) used to fit their data. Although Plack and Oxenham (1998) used a different value for w in the 1998 study, and Oxenham (2001) found better fits after allowing Tb1, Tb2, and w to vary, we felt the fit to our data was adequate using the original temporal window parameters listed above.

Buus, S., and Florentine, M. (**1991**). ''Psychometric functions for level discrimination,'' J. Acoust. Soc. Am. **90**, 1371–1380.

Dai, H. (**1995**). ''On measuring psychometric functions: a comparison of the constant-stimulus and adaptive up-down methods,'' J. Acoust. Soc. Am. **98**, 3135–3139.

Egan, J. P., Lindner, W. A., and McFadden, D. (**1969**). ''Masking-level differences and the form of the psychometric function,'' Percept. Psychophys. **6**, 209–215.

Fastl, H. (**1979**). ''Temporal masking effects: III. Pure tone masker,'' Acustica **43**, 282–294.

Glasberg, B. R., and Moore, B. C. J. (**2000**). ''Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise,'' J. Acoust. Soc. Am. **108**, 2318–2328.

Guilford, J. P. (**1965**). *Fundamental Statistics in Psychology and Education* (McGraw-Hill, New York).

Jesteadt, W., Bacon, S. P., and Lehman, J. R. (**1982**). ''Forward masking as a function of frequency, masker level, and signal delay,'' J. Acoust. Soc. Am. **71**, 950–962.

Kidd, Jr., G., and Feth, L. L. (**1981**). ''Patterns of residual masking,'' Hear. Res. **5**, 49–67.

Leek, M. R. (**2001**). ''Adaptive procedures in psychophysical research,'' Percept. Psychophys. **63**, 1279–1292.

Leek, M. R., Hanna, T. E., and Marshall, L. (**1992**). ''Estimation of psychometric functions from adaptive tracking procedures,'' Percept. Psychophys. **51**, 247–256.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49(2, Part 2)**, 467–477.

Luscher, E., and Zwislocki, J. (**1947**). ''The decay of sensation and the remainder of adaptation after short pure-tone impulses on the ear,'' Acta Oto-Laryngol. **35**, 428–455.

Moore, B. C., Peters, R. W., and Glasberg, B. R. (**1999**). ''Effects of frequency and duration on psychometric functions for detection of increments and decrements in sinusoids in noise,'' J. Acoust. Soc. Am. **106**, 3539–3552.

Munson, W. A., and Gardner, M. B. (**1950**). ''Loudness patterns—a new approach,'' J. Acoust. Soc. Am. **22**, 177–190.

Neff, D. L. (**1985**). ''Stimulus parameters governing confusion effects in forward masking,'' J. Acoust. Soc. Am. **78**(6), 1966–1976.

Neff, D. L. (**1986**). ''Confusion effects with sinusoidal and narrow-band noise forward maskers,'' J. Acoust. Soc. Am. **79**(5), 1519–1529.

Nelson, D. A., Chargo, S. J., Kopun, J. G., and Freyman, R. L. (**1990**). ''Effects of stimulus level on forward-masked psychophysical tuning curves in quiet and in noise,'' J. Acoust. Soc. Am. **88**, 2143–2151.

Nelson, D. A., and Freyman, R. L. (**1987**). ''Temporal resolution in sensorineural hearing-impaired listeners,'' J. Acoust. Soc. Am. **81**, 709–720.

Nelson, D. A., and Freyman, R. L. (**1984**). ''Broadened forward-masked tuning curves from intense masking tones: delay-time and probe-level manipulations,'' J. Acoust. Soc. Am. **75**, 1570–1577.

Nelson, D. A., Schroder, A. C., and Wojtczak, M. (**2001**). ''A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **110**, 2045–2064.

Nizami, L., and Schneider, B. A. (**1999**). ''The fine structure of the recovering auditory detection threshold,'' J. Acoust. Soc. Am. **106**, 1187–1190.

Oxenham, A. J. (**2001**). ''Forward masking: adaptation or integration?'' J. Acoust. Soc. Am. **109**, 732–741.

Oxenham, A. J., and Moore, B. C. J. (**1994**). ''Modeling the additivity of nonsimultaneous masking,'' Hear. Res. **80**, 105–118.

Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.

Oxenham, A. J., and Plack, C. J. (**2000**). ''Effects of masker frequency and duration in forward masking: further evidence for the influence of peripheral nonlinearity,'' Hear. Res. **150**, 258–266.

Plack, C. J., and Oxenham, A. J. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**, 1598–1608.

Rhode, W. S. (**1971**). ''Observations of the vibration of the basilar membrane in squirrel monkeys using the Mossbauer technique,'' J. Acoust. Soc. Am. **49**, 1218–1231.

Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). ''Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and response phases,'' J. Acoust. Soc. Am. **80**, 1364–1374.

Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.

Spiegel, M. F., and Green, D. M. (**1981**). ''Two procedures for estimating internal noise,'' J. Acoust. Soc. Am. **70**, 69–73.

Watson, C. S., Franks, J. R., and Hood, D. C. (**1972**). ''Detection of tones in the absence of external masking noise. I. Effects of signal intensity and signal frequency,'' J. Acoust. Soc. Am. **52**, 633–643.

Werner, L. A. (**1999**). ''Forward masking among infant and adult listeners,'' J. Acoust. Soc. Am. **105**(4), 2445–2453.

Widin, G. P., and Viemeister, N. F. (**1979**). ''Intensive and temporal effects in pure-tone forward masking,'' J. Acoust. Soc. Am. **66**, 388–395.

Yates, G. K., Winter, I. M., and Robertson, D. (**1990**). ''Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range,'' Hear. Res. **45**, 203–219.

Zwislocki, J., Pirodda, E., and Rubin, H. (**1959**). ''On some poststimulatory effects at the threshold of audibility,'' J. Acoust. Soc. Am. **31**, 9–14.

Psychophysical evidence for auditory compression at low characteristic frequencies

Christopher J. Plack[a] and Vit Drga
Department of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, England

(Received 2 May 2002; revised 23 November 2002; accepted 25 November 2002)

Psychophysical estimates of compression often assume that the basilar-membrane response to frequencies well below characteristic frequency (CF) is linear. Two techniques for estimating compression are described here that do not depend on this assumption at low CFs. In experiment 1, growth of forward masking was measured for both on- and off-frequency pure-tone maskers for pure-tone signals at 250, 500, and 4000 Hz. The on- and off-frequency masking functions at 250 and 500 Hz were just as shallow as the on-frequency masking function at 4000 Hz. In experiment 2, the forward masker level required to mask a fixed low-level signal was measured as a function of the masker–signal interval. The slopes of these functions did not differ between signal frequencies of 250 and 4000 Hz for the on-frequency maskers. At 250 Hz, the slope for the 150-Hz masker was almost as steep as that for the on-frequency masker, whereas at 4000 Hz the slope for the 2400-Hz masker was much shallower than that for the on-frequency masker. The results suggest that there is substantial compression, of around 0.2–0.3 dB/dB, at low CFs in the human auditory system. Furthermore, the results suggest that at low CFs compression does not vary greatly with stimulation frequency relative to CF. [DOI: 10.1121/1.1538247]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Mk [NFV]

I. INTRODUCTION

It is now well established that the response of the basilar membrane (BM) is highly compressive in the basal region of the cochlea, the region that is most sensitive to high frequencies (Murugasu and Russell, 1995; Rhode and Recio, 2000; Rhode and Robles, 1974; Robles *et al.*, 1986; Ruggero *et al.*, 1997; Russell and Nilsen, 1997). Much less is known about the response of the BM in the apical region of the cochlea, the region that is most sensitive to low frequencies. Direct measurements of BM vibration in the apical region *in vivo* are complicated by the difficulty of the surgical procedure, and by the obstruction provided by Reissner's membrane. Rhode and Cooper (1996) made BM and tectorial membrane (TM) displacement measurements in the apical turn of the chinchilla cochlea using laser interferometry, for places with characteristic frequencies (CFs) ranging from 500 to 800 Hz. They found a maximum compression of around 0.5 dB/dB, much less than the value of around 0.2 dB/dB typically estimated from the basal measurements. However, the technique involved opening Reissner's membrane in order to place microbeads onto the BM or the TM. This may have disrupted the ionic balance of the scala media, thereby damaging the physiologically vulnerable mechanism that underlies the nonlinear behavior. On the other hand, Zinn *et al.* (2000) reflected the measuring laser directly from Hensen cells in the guinea-pig cochlea, without damaging Reissner's membrane. They reported a nearly linear response for CFs of around 400 Hz.

Cooper and Yates (1994) estimated BM compression from auditory nerve rate-level functions recorded from guinea pigs. If it is assumed that the BM response to a tone well below CF is roughly linear (Ruggero *et al.*, 1997; Russell and Nilsen, 1997), then the BM response to a tone at CF can be derived from a comparison of the rate-level functions to tones at and below CF. For fibers with CFs between 1.5 and 3.6 kHz, the estimated compression was 0.5 dB/dB, compared to a maximum compression of 0.13 dB/dB for fibers with CFs above 4 kHz. Consistent with this, the threshold elevation in cat auditory nerve fibers produced by furosemide, which is thought to impair the active mechanism in the cochlea, is less for fibers with CFs below 3 kHz than for higher CFs (Sewell, 1984).

Over the last few years, psychophysical techniques have been used to make inferences about the response of the human cochlea. Some of these techniques involve comparing the growth of masking for forward maskers at or below the signal frequency (Moore *et al.*, 1999; Nelson *et al.*, 2001; Oxenham and Plack, 1997). Forward masking is used to avoid nonlinear interactions (e.g., suppression) between the masker and the signal in the cochlea. Following Yates *et al.* (1990) and Cooper and Yates (1994), these techniques assume that the BM response to a tone well below CF is roughly linear. The ratio of the slopes of the masking functions for on- and off-frequency pure-tone maskers provides an estimate of the BM compression to a tone at CF. Alternatively, it is sometimes assumed that the on-frequency masking function is linear (when the masker and signal have similar levels), and the BM response is derived from the off-frequency masking data alone. Plack and Oxenham (2000) presented a masker and signal alternately in a pulsation threshold design (Houtgast, 1972). The masker frequency was set to 0.6 times the signal frequency. Based on the assumption of below-CF linearity, they estimated BM compression for a range of signal frequencies. The compression ranged from 0.76 dB/dB at 250 Hz to around 0.35 dB/dB for

[a]Electronic mail: cplack@essex.ac.uk

 0001-4966/2003/113(3)/1574/13/$19.00

frequencies of 1000 Hz and above. Although the high-frequency compression was less than expected, the results seem to confirm that the cochlea is much less compressive in the low-frequency region. Hicks and Bacon (1999) explored the effects of frequency using three different psychophysical estimates of nonlinearity: frequency selectivity as a function of level, two-tone suppression, and the growth of forward masking for a masker below the signal frequency. Once again, all the measures showed a progressive decrease in nonlinearity from 3000 to 750 Hz, with no evidence of nonlinear processing at 375 Hz.

Although there seems to be a consensus among these studies that there is less compression in the apical region, some of the techniques depend on a crucial assumption: that the BM response to tones below CF is linear. In the techniques of Cooper and Yates (1994) and Oxenham and Plack (2000), the derived slopes will only differ from unity if there is *differential* compression between the on- and off-frequency tones. If both the masker and the signal are compressed equally, then the two effects should cancel out, and the derived response will be linear (1 dB/dB). The two additional psychophysical techniques used by Hicks and Bacon (1999) may also depend on this assumption. Frequency selectivity will change as a function of level only if the growth of excitation (and hence the compression) differs between frequency regions relative to CF. Although psychophysical two-tone suppression is more complex, it is possible to imagine situations in which there is a great deal of compression without any suppression. For example, if compression is applied after filtering (and is therefore in addition not frequency selective), then the nominal suppressor can only add to the overall output of the channel, not reduce it.

In other words, the psychophysical results are consistent with the possibility that the human BM response function is just as compressive in low-CF regions as in high-CF regions, *if* compression in the apical turn extends over a wide range of frequencies relative to CF. In fact, this is exactly what has been found in the chinchilla. The apical BM displacement measurements of Rhode and Cooper (1996) suggest that the compression may indeed be less frequency selective in the apex than in the base, with a broad band of frequencies affected relative to CF. This finding casts doubt on the validity of many of the psychophysical procedures used to estimate apical compression.

The aim of the present study was to investigate the frequency dependence of compression in human listeners using two psychophysical techniques that do not depend on below-CF linearity. The first technique is a measurement of the growth of forward masking. It has been known for many years that the growth of forward masking is nonlinear (Jesteadt *et al.*, 1982; Moore and Glasberg, 1983; Munson and Gardner, 1950). A given increase in masker level causes a much smaller increase in the signal level at threshold, if the signal level is low (below about 40 dB SPL). When both the signal level and the masker level are high, the growth of masking is roughly linear (Munson and Gardner, 1950; Oxenham and Plack, 1997; Plack and Oxenham, 1998). The shallow growth for low signal levels can be explained on the basis that the response of the BM to a tone at CF is linear at low levels and compressive at higher levels (Murugasu and Russell, 1995; Ruggero *et al.*, 1997). If the signal is in the low-level, linear region, then a given increase in signal level (in dB) will produce the same increase in BM vibration (in dB). If the masker is at a higher level, and is therefore in the compressive region of the BM function, then a given increase in masker level will produce a smaller increase in BM vibration. It follows that if the signal-to-masker ratio in terms of BM vibration is to remain constant, physical masker level has to grow more rapidly than physical signal level, leading to a shallow slope.

A prediction of this hypothesis is that if the signal level at threshold is made high enough so that it too falls within the compressive region of the BM, then both the masker and the signal response should grow at the same rate and the growth of masking should be *linear*. Plack and Oxenham (1998) tested this prediction by using short masker–signal intervals and brief signals to increase the signal level at threshold. Their results confirmed the prediction. As signal threshold reached a level of around 35 dB SPL, the growth function steepened dramatically, reaching unity at high levels. Plack and Oxenham showed further that the slope of the growth-of-masking function depended on signal level (rather than on masker level or on masker-signal interval) and that the overall form of the masking function was *quantitatively* consistent with the form of the BM response function. Further evidence that the forward masking growth function is related to cochlear compression comes from studies of forward masking in listeners with sensorineural hearing loss. Physiological models have indicated that sensorineural hearing loss is associated with a linearization of the BM response (Ruggero and Rich, 1991; Ruggero *et al.*, 1997). That is, animals with a moderate to severe cochlear hearing loss (produced in the laboratory) tend to show a linear BM response to a tone at CF. For human listeners with sensorineural hearing loss the growth of forward masking is linear (Moore *et al.*, 1999; Oxenham and Moore, 1995, 1997; Oxenham and Plack, 1997). This suggests, first, that the nonlinear growth of forward masking in normal listeners is the result of cochlear nonlinearity, and, second, that subsequent to the cochlea forward masking may be regarded as a *linear* process.

If the hypothesis is correct, then the growth of forward masking for low signal levels may provide an estimate of the BM compression in response to the masker, at the place with a CF equal to the signal frequency. The technique may be useful, therefore, in estimating the variation in BM compression with CF. In previous studies, little variation in the growth of forward masking with frequency has been observed (Harris and Rawnsley, 1953; Harris *et al.*, 1951; Jesteadt *et al.*, 1982; Lüscher and Zwislocki, 1949; Rawnsley and Harris, 1952), with shallow slopes across all frequencies, even down to 125 Hz in the experiment of Jesteadt *et al.*, for example. This is consistent with a compressive BM response at low CFs, although it has been suggested that the form of these functions may not be attributable solely to BM compression (Hicks and Bacon, 1999). The present study explores growth of forward masking functions in detail, including a range of masker frequencies relative to signal

frequency, at low and high signal frequencies. If the shallow masking functions are a result of BM compression, and if the compression is not frequency selective at low frequencies, then the masking functions should have the same slope for on- and off-frequency maskers. This is in contrast to forward-masking functions for high-frequency (e.g., 4 kHz), low-level signals, which show a shallow slope for on-frequency maskers and a linear slope for off-frequency maskers (Oxenham and Plack, 2000).

The second technique is a variation of that developed recently by Nelson *et al.* (2001). In this technique, the signal is fixed at a low level and the gap between the signal and the forward masker varied. For each gap, the masker level needed to mask the signal is measured for various masker frequencies. As the gap is increased, the masker level needed also increases, although the growth functions (masker level versus gap) for maskers close to the signal frequency are generally steeper than those for maskers below the signal frequency. (Nelson *et al.* refer to the resultant curves as *iso-response* temporal masking curves, or ''TMCs.'') The ratio of the slopes for on- and off-frequency maskers provides an estimate of the BM response, assuming that the response to the off-frequency masker is linear. At low frequencies, of course, this assumption may not be valid. However, if it is assumed that after the cochlea the decay of forward masking is the same across frequency, or possibly that the equivalent rectangular duration of the temporal window is constant across frequency (Plack and Moore, 1990), then the off-frequency masking functions at high frequencies may be compared to the on or off-frequency masking functions at low frequencies to obtain an estimate of the BM response functions at low CFs.

II. EXPERIMENT 1: GROWTH OF FORWARD MASKING

A. Stimuli

Three conditions were run, two low-frequency conditions and one high-frequency condition. In the first condition, the sinusoidal signal was set at $f_s = 250$ Hz and the sinusoidal masker at $f_m = 100$, 150, 250, or 400 Hz. In the second condition, the signal was set at $f_s = 500$ Hz and the masker at $f_m = 300$ or 500 Hz. In the third condition, the signal was set at $f_s = 4000$ Hz and the masker at $f_m = 2400$ or 4000 Hz. In all conditions maskers were presented at fixed levels of 35, 45, 55, 65, 75, 85, and 95 dB SPL and the signal level varied adaptively, starting at levels about 20 dB above masked threshold.

Temporal envelopes were the same for all conditions. The signal envelope consisted of 16-ms raised-cosine ramps with 0-ms steady state, giving an absolute duration (envelope 0 V to 0 V) of 32 ms. The masker envelope consisted of 16-ms raised-cosine ramps at each end of a 40-ms steady-state period, giving an absolute duration of 72 ms. Maskers always preceded signals. The masker–signal gap (between the 0-V points at the end of the masker and the start of the signal) was either 0 or 10 ms in the 250- and 4000-Hz conditions, and 0 ms in the 500-Hz condition.

The experiment was controlled by custom-made software from a PC workstation located outside a double-walled sound-attenuating booth. All stimuli were generated digitally with 32-bit resolution and were output by an RME Digi96/8 PAD 24-bit soundcard set at a clocking rate of 48 kHz. The soundcard includes an anti-aliasing filter. The output of the soundcard was fed from the headphone output of the soundcard via a patch-panel in the sound booth wall to Sennheiser HD 580 circumaural headphones without filtering or amplification. Stimuli were presented to the right ear. Each listener sat in the booth and decisions were recorded via a computer keyboard. Listeners viewed a computer monitor through a window in the sound booth. Lights on the monitor display flashed on and off concurrently with each stimulus presentation and provided feedback at the end of each trial.

B. Procedure

A two-interval forced-choice forward masking experiment was run using an interstimulus interval of 500 ms. The masker was presented in both intervals. The signal was presented following the masker in only one of the two intervals chosen at random with probability 0.5. Masker level and frequency remained fixed within each block of trials. Signal level was varied adaptively using a two-down one-up adaptive procedure to obtain the signal level needed to achieve 70.7% correct (Levitt, 1971). The signal level step size was 4 dB for the first 4 turnpoints, which reduced to 2 dB for 12 subsequent turnpoints. The mean of the last 12 turnpoints was taken as the threshold estimate for each block of trials. If the standard deviation of the turnpoints was greater than 6 dB, the estimate was discarded and the trial-block was later repeated. At 250 and 4000 Hz, five estimates were averaged and taken as the threshold for each combination of parameters. At 500 Hz, ten estimates were averaged.

Listeners ran in 1–2-h sessions and took breaks as needed. Per replication, there were 56 blocks in the 250-Hz condition (7 masker levels×4 masker frequencies×2 masker–signal gaps), 14 blocks in the 500-Hz condition (7 masker levels×2 masker frequencies×1 masker–signal gap), and 28 blocks in the 4000-Hz condition (7 masker levels×2 masker frequencies×2 masker–signal gaps). Blocks were presented in randomized order within each replication. The 250- and 4000-Hz replications were alternated until five of each had been completed. The 500-Hz replications were run later in separate sessions.

C. Listeners

Three normal-hearing listeners took part in the experiment. Listeners EK and PP were college students paid an hourly wage for their participation, and author VD was also a listener. EK and VD were experienced listeners whereas PP was a naïve listener. Absolute thresholds for the signal (for VD, EK, and PP, respectively) were 10.6, 22.4, and 19.9 dB SPL at 250 Hz;[1] 7.2, 10.5, and 14.0 dB SPL at 500 Hz; and 0.0, 8.7, and 1.7 dB SPL at 4000 Hz. Each listener received at least three hours practice before data collection began.

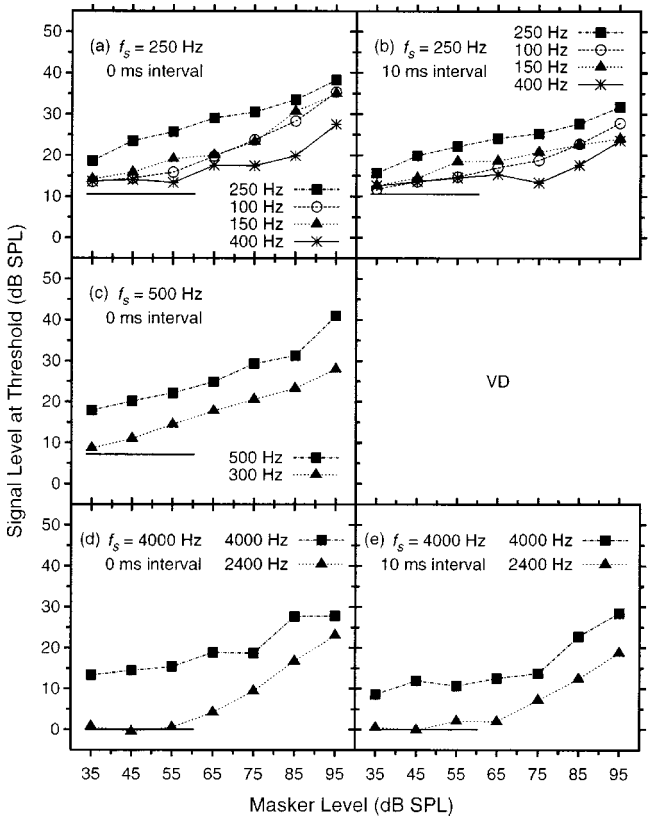


FIG. 1. The results of experiment 1 for listener VD, showing signal level at threshold as a function of masker level. Data for the 250-Hz signal are shown in panels (a) and (b), for the 500-Hz signal in panel (c), and for the 4000-Hz signal in panels (d) and (e). Left-hand panels show data for the 0-ms masker–signal interval. Right-hand panels show data for the 10-ms masker–signal interval. Masker frequencies are listed in each panel next to their respective data symbols. Solid squares always represent on-frequency masking at each signal frequency. Horizontal lines indicate the absolute threshold for the signal at each frequency.

D. Results, analysis, and discussion

The individual data are presented in Figs. 1–3. The figures show signal level at threshold as a function of masker level. Each panel shows the data for a given signal frequency and a given masker–signal interval. The different curves show the results for different masker frequencies. The horizontal lines in each panel show the absolute thresholds for the signal.

The broad pattern of results was similar for the three listeners. Consider first the results for the 250-Hz signal. All the curves show a shallow rate of growth of signal threshold with masker level. For the 0-ms gap the curves for the low-frequency maskers ($f_m = 100$ and 150 Hz) are slightly steeper than the on-frequency curve ($f_m = 250$ Hz), over the range of masker levels from 55 to 95 dB SPL. For the 10-ms gap the slopes for all the masker frequencies are similar. The results at 500 Hz show similarly shallow slopes for the on- and off-frequency maskers, except that, for two of the listeners (VD and EK), there is a sudden increase in threshold for the highest masker level in the on-frequency condition. Similarly for the 4000-Hz conditions, the slopes for the on-frequency masker tend to steepen slightly at the highest masker levels in some cases. The slopes for the off-frequency masker are very shallow for low masker levels (probably because threshold is limited by the absolute threshold for the signal) but then steepen, converging on the slopes for the on-frequency masker. For a given *signal* level at 4000 Hz, the slopes for the off-frequency masker are generally steeper than the slopes for the on-frequency masker, a pattern that is not seen for the two lower signal frequencies.

In their studies of forward masking, Oxenham and Moore (1994) and Plack and Oxenham (1998) interpreted the on-frequency growth-of-masking slopes as a reflection of the characteristics of the BM response function to a tone at CF. When the signal is at a low level it is passed by the BM roughly linearly, but the more-intense masker is compressed. It follows that if the signal-to-masker ratio in terms of BM vibration remains constant, the slope of the growth-of-masking curve should be numerically equal to the compression applied to the masker (in dB/dB). When the signal level enters the compressive portion of the response function the slopes steepen towards linearity, as both masker and signal are compressed. This may explain the increase in slope at high levels, shown particularly in the 500-Hz data. At high CFs the response to a masker much lower than CF will be roughly linear at all levels. When the signal level is low, therefore, both the masker and the signal will be passed linearly. This should result in a linear growth of masking (1 dB/dB). It appears, however, that the slope for a 2400-Hz masker and a 4000-Hz signal is less than 1 in the present data. A possible reason for this discrepancy is discussed in Sec. IV B.

The masking functions for the results averaged across the three listeners were subjected to straight-line, least-squares fits, considering only those signal levels greater than 5 dB SL. (It was considered that thresholds less than this would be determined more by the proximity of absolute threshold than by the level of the masker.) For two of the data sets (on-frequency 500 and 4000 Hz, 0-ms gap) where there appeared to be a distinct steeping of the curve at high levels in the mean data, two lines were fit, one for the shallow growth and one for the steeper growth. This was in addition to a fit covering all the data for the condition. The slopes of the straight-line fits are shown in Table I, together with the range of masker levels used for each fit. Following the reasoning of Plack and Oxenham, and with the exception perhaps of the two upper-segment fits, the slopes can be considered estimates of the BM compression applied to the masker for each data set.

Considering first the single fits to the on-frequency masker data, there is comparatively little variability across the three frequencies and two gaps, with values ranging from 0.2 dB/dB (250 Hz, 10 ms) to 0.34 dB/dB (500 Hz, 0 ms). All these values are within the range usually associated with the basal region of the cochlea. Dividing the data into two segments reduces the slope of the lower-level portion to 0.27 dB/dB at 500 Hz and 0.15 dB/dB at 4000 Hz, 10-ms gap. The slopes for the off-frequency maskers are also revealing. For $f_s = 4000$ and $f_m = 2400$ Hz, the slope is much steeper than the slope for the on-frequency masker, having a value of 0.62 dB/dB for the 0-ms gap and 0.54 dB/dB for the 10-ms gap. While these values are a fair way from the linearity expected by theory, they are considerably higher than

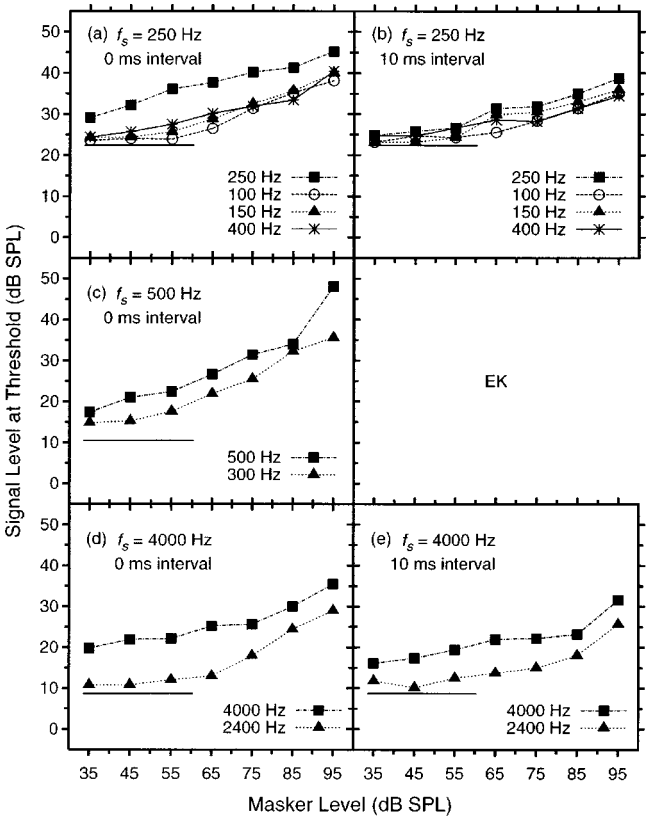


FIG. 2. As in Fig. 1 except showing the results for listener EK.

any of the slopes for the on-frequency masker for low signal levels. [Oxenham and Plack (1997) observed roughly linear growth of forward masking for low-level signals and low-frequency maskers, but they used high-pass noise to restrict off-frequency listening; see Sec. IV B.] In contrast, the slopes for the off-frequency maskers for 250 and 500 Hz from the present study are not dramatically different from the slopes for the on-frequency maskers. The biggest frequency effect is seen for the 250-Hz, 0-ms conditions, and even here the slopes for $f_m = 100$ and $f_m = 150$ Hz are only 0.37 dB/dB, compared to 0.24 dB/dB for the on-frequency condition. There is a slight tendency for the slopes to decrease as the gap is increased at 250 and 4000 Hz. This may partly reflect the closer proximity of the thresholds to absolute threshold for the 10-ms gap. Clearly, at absolute threshold itself the masker will have no effect on signal threshold and the slope will be 0.

Considered overall, and in the context of the explanation for forward-masking growth proposed by Plack and Oxenham (1998), the results of experiment 1 provide evidence for strong compression at low CFs that does not vary substantially with frequency relative to CF.

III. EXPERIMENT 2: TEMPORAL MASKING CURVES

A. Stimuli

Two conditions were run, a low-frequency and a high-frequency condition. Each condition involved both on-frequency masking ($f_m = f_s$) and off-frequency masking ($f_m = 0.6 f_s$). In the low-frequency condition, the sinusoidal signal was set at $f_s = 250$ Hz and the sinusoidal masker at $f_m = 150$ or 250 Hz. In the high-frequency condition, the signal was set at $f_s = 4000$ Hz and the masker at $f_m = 2400$ or 4000 Hz.

The temporal envelopes were the same as in experiment 1 except that a larger range of masker–signal gaps was used. The minimum gap used was always 0 ms. The maximum gap depended on each individual's performance in each condition, and extended to values where practice blocks indicated masker levels at threshold would lie above 90 dB SPL. For listeners VD, EK, and PP, gaps ranged out to 52, 44, and 24 ms, respectively, in the 250-Hz condition, and 70, 45, and 55 ms, respectively, in the 4000-Hz condition. For all listeners in the 250-Hz condition, and for EK in the 4000-Hz condition, the range of gaps was smaller for off-frequency masking than for on-frequency masking. For the high-frequency condition, gaps generally were set in multiples of 5 ms, but also in multiples of 2.5 ms over the range where a listener's on-frequency masker thresholds increased most rapidly as a function of gap. For the low-frequency condition, gaps were set in multiples of 4 ms for listeners VD and EK and in multiples of 2 ms for PP, who showed much steeper TMCs. The method of stimulus generation and presentation was the same as for experiment 1.

B. Procedure

The basic procedure was the same as in experiment 1, with the following exceptions. The signal was *fixed* at 6 dB above absolute threshold (6 dB SL, relative to the absolute threshold measured in experiment 1) and the masker level was varied adaptively using a two-up one-down adaptive procedure to obtain the masker level needed to achieve 70.7% correct (Levitt, 1971). The initial masker level was set about 20 dB below the estimated masked threshold (based on practice trials). The masker level step size was 4 dB for the first 4 turnpoints which reduced to 2 dB for the 12 subsequent turnpoints. The mean of the last 12 turnpoints was taken as the threshold estimate for each block of trials. If the standard deviation of the turnpoints was greater than 6 dB, the estimate was discarded and the block was later repeated. Data were also discarded and repeated for the occasional block in which the masker clipped at levels greater than or equal to 102 dB SPL.[2]

All the data for $f_s = 4000$ Hz were collected before the data for $f_s = 250$ Hz. Trials were run in sessions of 15–60 min, with listeners taking breaks as needed between sessions. The order of masker–signal gaps and on- and off-frequency maskers was randomized together across blocks of trials. In other words, within a replication, the block was selected randomly from among all the remaining masker–signal interval and frequency combinations. On account of the number of conditions, and the variable session length, a complete TMC was rarely completed within a session. The remaining conditions were completed in a random order in the subsequent session. Five replications were run in the high-frequency condition. Ten replications were run in the low-frequency condition due to greater variability in the data. The mean over five (or ten) estimates was used as the threshold for each combination of masker–signal gap and masker fre-

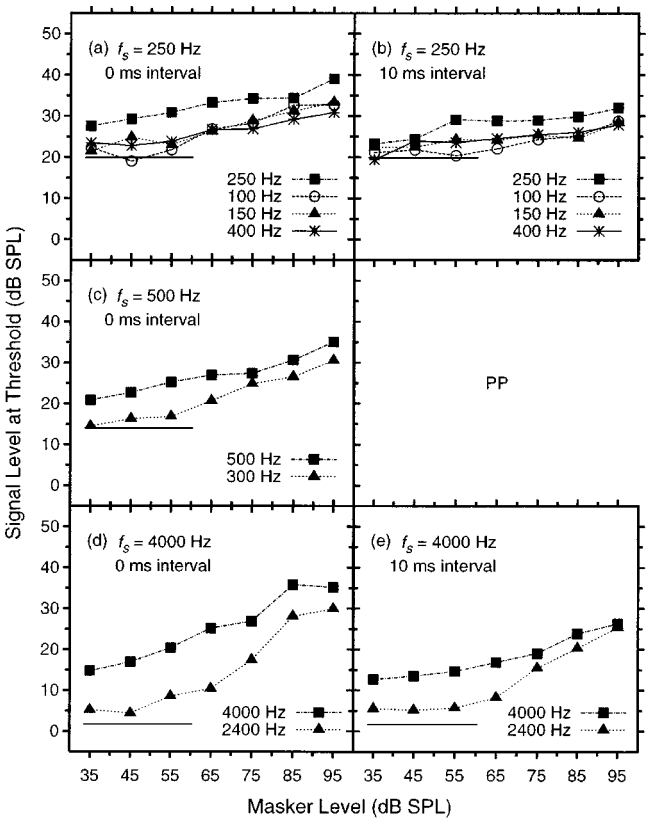


FIG. 3. As in Fig. 1 except showing the results for listener PP.

quency. No discernable practice effects were evident in the data.

The listeners were the same as in experiment 1. Each listener ran several sets of practice replications to determine the range of masker–signal gaps to use for the main data collection.

TABLE I. Slopes of the mean growth of masking functions for each masker frequency and masker–signal interval in experiment 1. On-frequency-masker functions for 500 and 4000 Hz are divided into two segments, roughly following the breakpoints seen in the mean data. Slopes are given for each segment as well as for the entire function, and were only calculated for data points lying 5 dB or more above mean absolute threshold, as indicated by the masker range.

Masker frequency, f_m (Hz)	0-ms gap		10-ms gap	
	Masker range (dB)	Slope (dB/dB)	Masker range (dB)	Slope (dB/dB)
100	65–95	0.37	75–95	0.34
150	65–95	0.37	65–95	0.17
250	35–95	0.24	45–95	0.20
400	65–95	0.26	65–95	0.20
300	55–95	0.37		
500	35–95	0.34		
500	35–85	0.27		
500	85–95	0.94		
2400	65–95	0.62	75–95	0.54
4000	35–95	0.29	35–95	0.25
4000	35–75	0.21	35–75	0.15
4000	75–95	0.45	75–95	0.52

C. Results, analysis, and discussion

The individual TMCs are presented in Fig. 4. Shown are the on- and off-frequency masker levels required to mask the low-level signal as a function of the silent interval between the masker and the signal. For $f_s = f_m = 4000$ Hz, for listeners VD and PP the masker grows relatively slowly for the shorter gaps, before steepening dramatically at the longer gaps (higher masker levels). The same pattern was observed by Nelson *et al.* (2001) and is thought to reflect BM compression. As the masker level reaches the compressive region of the response function, larger changes in masker level are required to produce the same change in BM excitation (and in the amount of forward masking), and hence the TMC steepens. Listener EK does not show a striking increase in slope for the on-frequency condition. For all three listeners, however, the on-frequency TMCs are considerably steeper than the off-frequency TMCs ($f_s = 4000$ Hz, $f_m = 2400$ Hz), which grow at an almost constant shallow rate with increases in gap. If it is assumed that the BM response to a below-CF masker is linear, then the linear growth of masker level with gap for the off-frequency TMCs implies that the decay of forward masking, *measured in units of BM velocity or intensity*, is an exponential process. This is consistent with recent incarnations of the temporal window model (Oxenham and Moore, 1994; Plack and Oxenham, 1998; Plack *et al.*, 2002; Wojtczak *et al.*, 2001).

For the 250-Hz signal frequency the slopes of the TMCs (on- or off-frequency) vary little with increasing gap. Again, this may be because the masker thresholds for the shortest gap were higher than at 4000 Hz. Importantly, however, the on-frequency TMCs ($f_s = f_m = 250$ Hz) for all three listeners appear to be just as steep as the on-frequency TMCs at 4000 Hz. Two possible reasons for this come to mind. Either temporal resolution is more acute at low frequencies than at high frequencies (so that the effective decay in masker excitation with time is less) and the compression is less, or temporal resolution does not vary much with frequency and there is just as much compression at low CFs as at high CFs. The fact that temporal resolution measured by gap detection appears not to vary significantly with frequency (Shailer and Moore, 1987) suggests that the explanation in terms of equivalent compression across CF is the more plausible. Furthermore, for listeners VD and PP the off-frequency TMCs ($f_s = 250$ Hz, $f_m = 150$ Hz) appear almost as steep as the on-frequency TMCs. In accord with experiment 1, this suggests that for these two listeners at least compression does not vary substantially as stimulation frequency is varied relative to CF. For listener EK this does not seem to be the case. This listener shows a steeper TMC for the on-frequency as compared to the off-frequency masker.

Yates *et al.* (1990) derived BM response functions from auditory-nerve fiber rate-level functions by comparing on- and off-CF recordings. Using the same reasoning, Nelson *et al.* (2001) showed how it is possible to estimate BM response functions from TMC data. Briefly, it is assumed first that the BM excitation produced by the masker at the place tuned to the signal is *constant* for a given masker-signal gap at masked threshold. Second, it is assumed that the growth in BM excitation with masker level is *linear* for a masker well

below the signal frequency. It follows that, give or take an additive constant, the physical level of the low-frequency masker needed to mask the signal is an estimate of the BM excitation required. Hence if, for each gap, the off-frequency masker level at threshold is plotted against the on-frequency masker level at threshold, the curve describes the BM excitation produced by the on-frequency masker as a function of level. In other words, the curve represents the BM response function to a tone at CF.

The problem with applying this technique to the present data is that there seems to be evidence from both experiments that at low CFs the response to a below-CF masker is not linear. However, if it is assumed that temporal resolution does not vary with CF, then the TMC for the 2400-Hz masker can be used as a linear reference for all CFs. To do this, a straight-line fit was made to the 2400-Hz TMC for each listener independently. A projected value could then be read off to provide the *y*-axis reading (i.e., excitation estimate) for each masker level and gap. Response functions were derived in this way for both masker frequencies for the 250-Hz signal, and for the on-frequency masker for the 4000-Hz signal. The results are shown in Fig. 5.

To obtain an estimate of the slopes of the response functions, linear least-squares fits and also third-order polynomial least-squares fits were made to these data (Nelson *et al.*, 2001). The best fitting polynomials are shown by the continuous lines on Fig. 5. The first derivative of each polynomial provides an estimate of the slope of the response function, and hence the compression, as a function of input level. These are shown in Fig. 6. Slopes of the linear fits and minimum slopes of the cubic fits are given in Table II.

Slope estimates from the polynomial fits are in agreement with the estimates based on linear fits, but the minimum slopes are lower by a factor of $\frac{2}{3}$, broadly speaking. (This just reflects the ability of the cubic form to better approximate the data than a linear fit.) At 250 Hz, slopes range from 0.14 to 0.23, again indicating substantial compression at low frequencies.

It should be noted that a recent psychophysical study measured the variation in frequency selectivity with CF and level (Lopez-Poveda *et al.*, 2003). The technique based on TMCs was similar to that used in experiment 2, and the conclusions regarding the variation in compression with CF were consistent with those described here.

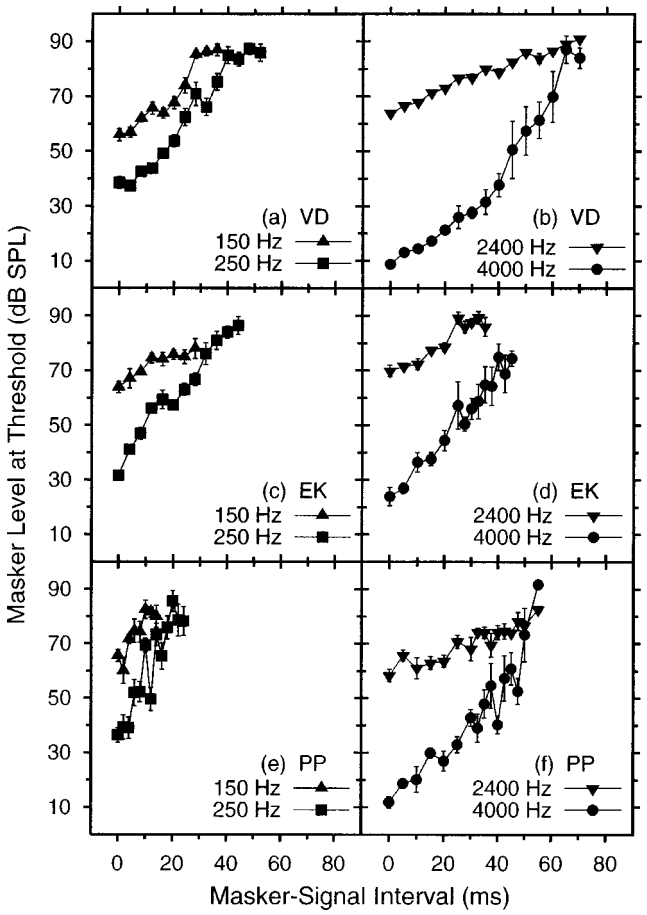


FIG. 4. Temporal masking curves (TMCs) for each of the three listeners in experiment 2, VD, EK, and PP, showing masker level at threshold as a function of masker–signal interval. Left-hand panels show TMCs for the 250-Hz signal. Right-hand panels show TMCs for the 4000-Hz signal. Masker frequencies are listed in each panel next to their respective data symbols. Error bars show ±1 standard error.

IV. GENERAL DISCUSSION

A. Analysis with the temporal window model

The temporal window model is a model of masking and of temporal resolution (Moore *et al.*, 1988; Oxenham and Moore, 1994; Plack and Moore, 1990; Plack and Oxenham, 1998; Plack *et al.*, 2002; Wojtczak *et al.*, 2001). The model assumes that the internal representation of a stimulus is smoothed in time by the action of a sliding intensity integrator, or temporal window. An assumption of the model is that temporal masking, both forward and backward, is a result of this integration process: The signal is masked because the temporal window centered on the time of signal presentation integrates masker energy for times before and after. The temporal window model was fit to the data from the two experiments for two reasons: first, to determine whether the results are consistent with a single function representing the decay of internal excitation (i.e., the temporal window itself), and, second, to provide an alternative measure of compression across frequency that takes into account, among other things, the proximity of some of the masked thresholds to absolute threshold in experiment 1. The model does this by incorporating an internal noise floor that limits the detectability of low-level signals.

The characteristics of the temporal window model used in the present article have been described in detail elsewhere (Plack and Oxenham, 1998), and only a brief description is provided here. The input stimulus is subjected to a ''broken-stick'' nonlinearity of the form:

$$L_{\text{out}} = L_{\text{in}}, \quad L_{\text{in}} < B, \tag{1}$$

$$L_{\text{out}} = cL_{\text{in}} + (1-c)B, \quad L_{\text{in}} \geqslant B, \tag{2}$$

where L_{in} (in dB SPL) and L_{out} (in dB) are the input and output levels, respectively, c is the compression exponent, and B is the breakpoint in the function (in dB SPL). In order to simplify the analysis, the lower slope of the function [Eq. (1)] is assumed to be linear. The output of the nonlinear

stage, in units of amplitude, is squared before being passed through the temporal window. The weighting function, $W(t)$, for the temporal window is defined by

$$W(t)=(1-w)\exp(t/Tb_1)+w\exp(t/Tb_2), \quad t<0, \tag{3}$$

$$W(t)=\exp(-t/Ta), \quad t\geqslant 0, \tag{4}$$

where t is time relative to the window peak in ms, and the parameter values are Tb_1, Tb_2, w, and Ta. The leading slope of the window (that determines forward masking) consists of two exponentials, one with time constant Tb_1 that determines the shape of the tip of the window, and one with time constant Tb_2 that determines the skirt of the window. The parameter w determines the relative weight given to these two functions. The equivalent rectangular duration (ERD) of the window is given by

$$\text{ERD}=(1-w)Tb_1+wTb_2+Ta. \tag{5}$$

To find predictions for a set of temporal masking data, the masker alone is passed through the model and then the masker plus signal is passed through the model. Detection is assumed to be based on the ratio of the signal-plus-masker at the output of the model to the masker alone at the output of the model, expressed in dB (in other words, a measure of ΔL), at the time that this ratio reaches its maximum value. To reduce the computational load, for forward masking the ''center time'' of the temporal window is assumed to be at the start of the offset ramp of the signal.

One of the parameters of the model is k, the value of the decision variable at threshold. In predicting a single data point, the model varies the dependent variable (e.g., signal threshold), repeating the simulation until the value of the decision variable is equal to k. The value of the dependent variable required is taken as the prediction of the model. This process is repeated for the entire data set. For each value of k a constant is added to the output of the temporal window so that, in the absence of the masker, the prediction of the model is equal to the absolute threshold determined experimentally. Effectively, the model simulates absolute threshold by assuming that there is an internal noise floor that contributes to the masking of signals close to absolute threshold.

Since the form of the TMC for the 2400-Hz data set should be determined by the shape of the temporal window only (assuming that the BM responds linearly to this frequency at the 4000-Hz place), it was decided to use these data to derive the shape of the temporal window for use in the simulations. The value of c was set to 1 (linear response) and the model was fit to the mean TMC data at 2400 Hz, varying Tb_1, Tb_2, w, and k (using the ''fminsearch'' routine in MATLAB) to obtain the values that minimized the squared deviations of the predicted thresholds from the measured thresholds. The best-fitting parameter values were 8.34, 30.29, and 0.000 73, for Tb_1, Tb_2, and w, respectively. (The best-fitting value of k, 0.000 22 dB, was determined mainly by the attenuation of the masker by the auditory filter, and it is not relevant to the rest of the analysis). Although these values differ from those used by Plack and Oxenham (4.0, 29.0, and 0.16, for Tb_1, Tb_2, and w, respectively), the ERDs of the windows are almost identical (11.50 ms in Plack and Oxenham compared to 11.85 ms here).

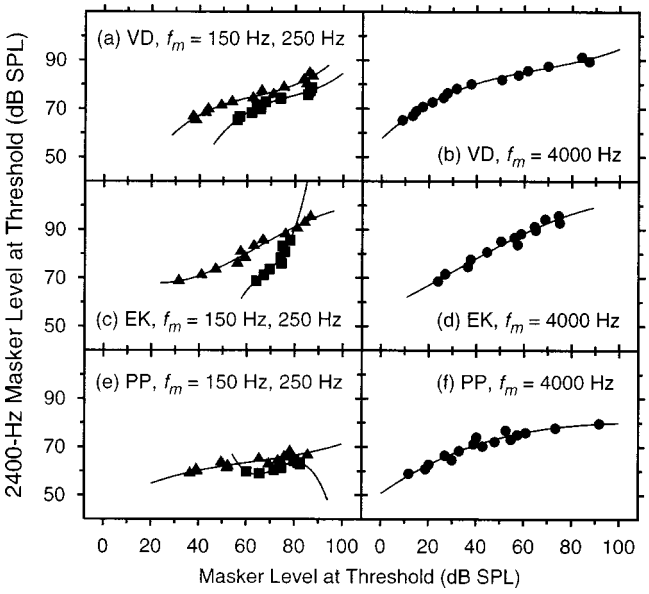


FIG. 5. Estimated response functions for each of the three listeners in experiment 2. Each function shows the 2400-Hz masker level at threshold versus the masker level at threshold for the masker frequency of interest, paired according to masker–signal interval, for $f_m=150$ Hz (squares), 250 Hz (triangles), and 4000 Hz (circles). Solid lines represent cubic fits to the functions. At each masker–signal interval, the 2400-Hz masker level for each listener is based on linear fits to the 2400-Hz temporal masking curves shown in Figs. 4(b), (d), and (f).

For the remaining fits the temporal window parameters were fixed. In the first phase, three fits were conducted: all the mean data for the 250-Hz signal and 250-Hz masker from the two experiments; all the mean data for the 4000-Hz signal and 4000-Hz masker from the two experiments; and the mean data for the 500-Hz signal and 500-Hz masker from experiment 1. The best-fitting values of c, B, and k were obtained for each of the three on-frequency data sets. In the second phase, three further fits were made: all the mean data for the 250-Hz signal and 150-Hz masker from the two experiments; all the mean data for the 4000-Hz signal and 2400-Hz masker from the two experiments; and the mean data for the 500-Hz signal and 300-Hz masker from experiment 1. To fit the off-frequency data, assumptions had to be made about the variation in nonlinearity with frequency. For the 250/150- and 500/300-Hz fits, it was assumed that all the parameters (including k) were the same as those derived in the first phase from the corresponding on-frequency fits (250/250 and 500/500 Hz, respectively). In particular, the nonlinearity for the off-frequency fits was identical to that for the on-frequency fits, consistent with the hypothesis that there is no variation in compression with frequency relative to CF. It was also assumed, however, that the off-frequency maskers were filtered prior to the nonlinearity. In the fitting procedure, these maskers were subject to an attenuation of A dB to reflect the filtering. For the 4000/2400-Hz fit, the value of c was set to 1, to reflect linear processing of below-CF stimuli, but again k was fixed at the value derived from the on-frequency fit, and the masker level attenuated by A. In all

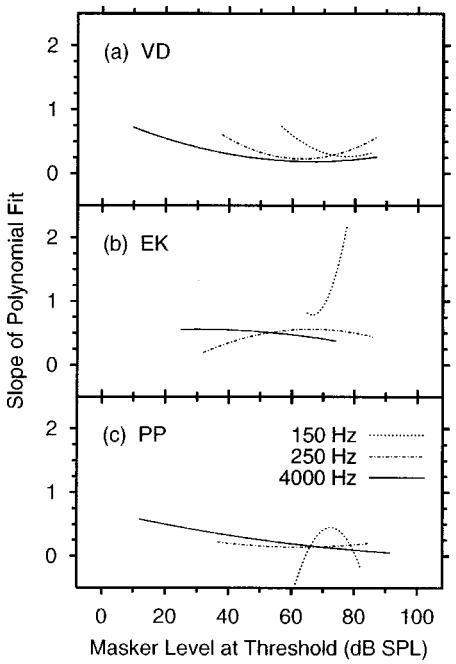


FIG. 6. Response function slope as a function of masker level, for masker frequencies $f_m = 150$ Hz (dotted lines), 250 Hz (dot-dash lines), and 4000 Hz (solid lines). These quadratic functions are the derivatives of the cubic functions shown in Fig. 5.

three off-frequency fits, *A* was the only parameter varied, and a single value of *A* was derived for each of the three fits.

The fits to the mean data are shown in Fig. 7 and the best-fitting parameters are shown in Table III. Considering that only a small number of parameters were varied in the fits, the model does a good job of accounting for the data, especially the on-frequency masking functions. It should be remembered that for a given signal frequency, the *same parameter values* were used for all the on-frequency data across the two experiments. It is reassuring to discover that the results are consistent between the two experiments. The same nonlinearity, the same temporal window, and the same detector efficiency can account for the results from the two different measurement techniques.

TABLE II. Slopes of estimated response functions at each masker frequency and for each listener in experiment 2. Linear fits were only calculated for masker thresholds greater than 40 dB SPL. [The minimum slope of the cubic fit to PP's data for 150 Hz is omitted since the slope becomes negative, as seen in Fig. 6(c).]

Listener	Masker frequency, f_m (Hz)	Slope of linear fit to response function	Minimum slope of cubic fit to response function
VD	4000	0.23	0.19
	250	0.31	0.23
	150	0.36	0.26
EK	4000	0.46	0.37
	250	0.52	0.19
	150	1.14	0.79
PP	4000	0.15	0.05
	250	0.15	0.14
	150	0.21	...

Another significant finding is that the off-frequency data at the two lower signal frequencies, 250 and 500 Hz, are broadly consistent with the nonlinearity derived from the on-frequency data, if it is assumed that the lower-frequency masker is subject to an attenuation (or filtering) before being passed through the nonlinear function. The assumption of a constant nonlinearity across frequency simplifies the fitting procedure considerably, because it is not necessary to fit separate nonlinearities for the masker and the signal when their frequencies differ. However, the mean signal thresholds for the off-frequency maskers in experiment 1 were all below the estimated breakpoints (*B*) in the broken-stick functions at 250 and 500 Hz (see Table III). It follows that, in the model, the representation of the signal will be unaffected by the compression exponent *c*. A new set of fits was therefore conducted on the off-frequency data at 250 and 500 Hz, allowing both the attenuation, *A*, and the compression exponent, *c*, to vary. All the other parameters were the same as for the on-frequency conditions. At 250 Hz, the best-fitting values of *c* and *A* were 0.21 and 26.26 dB, respectively, with a rms error of 2.03 dB. At 500 Hz, the best-fitting values of *c* and *A* were 0.29 and 24.88 dB, respectively, with a rms error of 1.35 dB. The compression values for the off-frequency maskers, measured independently, do not differ greatly from those for the on-frequency maskers in Table III. The fits suggest that the simplifying assumption of a constant nonlinearity across frequency is reasonable at low CFs.

For the 4000-Hz signal frequency it was assumed that the off-frequency masker grows linearly at the signal place. However, the assumption of below-CF linearity, while consistent with the results of experiment 2 at 4000 Hz, does not provide good fits to the off-frequency results of experiment 1 at 4000 Hz. The measured threshold values for the highest masker levels for both gap conditions are less than those predicted by the model. It is possible that the experimental values were affected by off-frequency listening, a process that is not taken into account in the current version of the model. This possibility will be explored in Sec. IV B.

The main finding of the modeling results is that the best-fitting value of the compression exponent, *c*, is almost constant across frequency at around 0.25. To estimate how sensitive the fits are to the value of *c*, the fits to the on-frequency data were repeated with the value of *c* fixed at either 0.75 or 1.25 times the best-fitting values. *B* and *k* were varied adaptively. At 250 Hz, the rms error increased from 1.64 dB (best fitting value of *c*) to 2.46 and 2.35 dB, for multipliers of 0.75 and 1.25, respectively. At 500 Hz, the rms error increased from 0.44 dB to 3.29 and 0.88 dB. At 4000 Hz, the rms error increased from 2.47 dB to 2.88 and 3.06 dB. Overall, therefore, the value of *c* appears to be fairly well constrained by the data, particularly at the two lower frequencies. The model fits support the hypothesis that compression does not vary greatly with CF in humans.

Finally, the transitions, or breakpoints, between the linear and nonlinear portions of the broken-stick function are similar at 500 and 4000 Hz (33 and 31 dB SPL) and are consistent with previous physiological and psychophysical

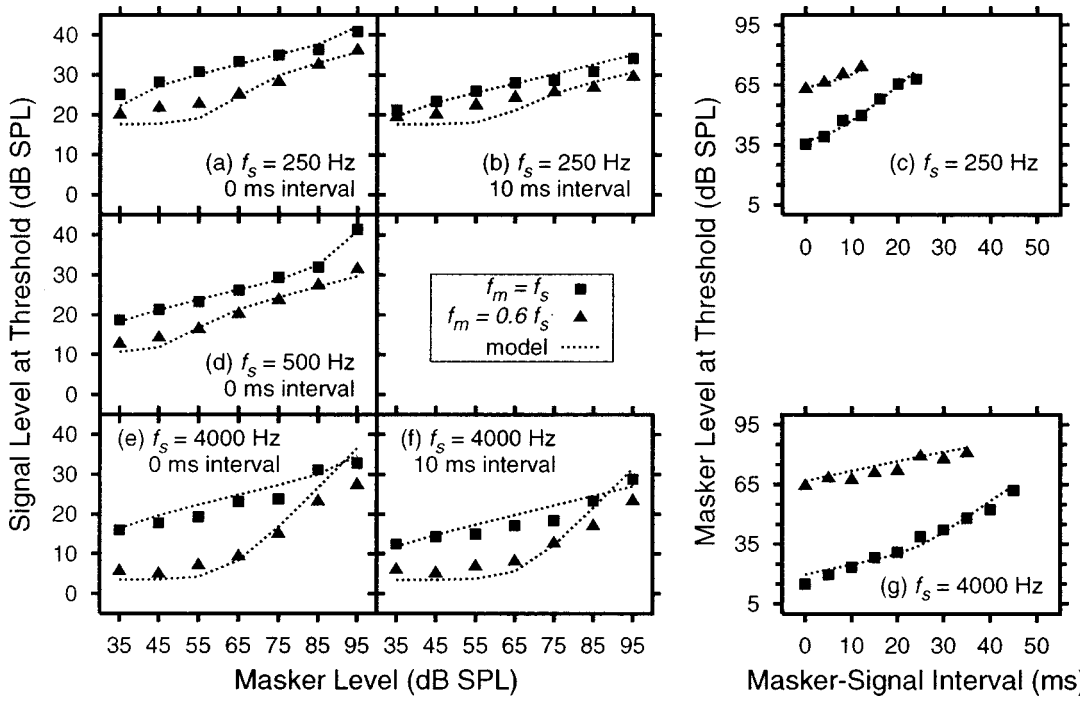


FIG. 7. The mean data from the two experiments together with the thresholds predicted by the temporal window model. The two columns on the left show the data from experiment 1. The right-hand column shows the data from experiment 2.

findings (e.g., Ruggero *et al.*, 1997; Plack and Oxenham, 1998). The estimated breakpoint at 250 Hz is slightly higher (41 dB SPL).

B. Off-frequency listening

It was noted in the discussion of the results from experiment 1 that the slopes of the masking functions for a 4000-Hz signal and a 2400-Hz masker, while greater than the slopes for the on-frequency masker, were less than would be expected if the auditory system were responding linearly to a below-CF stimulus. This was confirmed by the modeling results: The signal levels at threshold for the highest masker levels were less than the values predicted by the model. One possibility is that the results were affected by off-frequency listening (Johnson-Davies and Patterson, 1979; O'Loughlin and Moore, 1981). It has been shown that listeners do not always use the auditory filter centered on the signal frequency to detect the signal. For example, adding a narrow-band noise above the signal frequency that is not in itself sufficient to mask the signal will reduce the detectability of the signal in the presence of a low-frequency masker (Johnson-Davies and Patterson, 1979). It is argued that the narrow-band noise restricts the listening band, forcing listeners to use a region of the cochlea tuned close to the signal frequency. When the noise is not present, listeners may improve the signal-to-masker ratio by using an auditory filter centered above the signal frequency, or perhaps by combining information across a range of places in the cochlea.

As the signal level is increased, so the excitation produced by the signal spreads farther along the BM. The benefits of off-frequency listening should therefore increase with level. The increase in off-frequency listening with level may explain why the slopes of the masking functions were less than unity for the 4000-Hz signal and 2400-Hz masker in experiment 1. In their study using forward masking to estimate BM compression, Oxenham and Plack (1997) presented a high-pass noise when the masker was below the signal frequency to restrict off-frequency listening. In these conditions, the masking growth was *linear* (slope=1) when the signal level was low. Nelson *et al.* (2001) showed that, in the absence of the high-pass noise, the slopes of the functions relating masker level to signal level were much steeper (a relative increase in detectability with level), although the effect of the noise on signal detectability was small for signal levels below 40 dB SPL.

It is possible that the off-frequency results for the two lower frequencies (250 and 500 Hz) were also affected by off-frequency listening. However, the fact that these masking functions are much shallower than those for the 4000/2400-Hz conditions suggests that there is still substantial off-frequency compression at low CFs. Furthermore, the modeling results suggest that the off-frequency data at 250 Hz are broadly consistent between experiments 1 and 2. In experiment 2, the signal was fixed at a low level and off-frequency listening should not have had an effect on the slope of the TMC. Finally, it is unlikely that off-frequency listening had a significant effect on the on-frequency masking functions. For low-level signals and on-frequency maskers, the effect of restricting the listening bandwidth with noise is minimal (Johnson-Davies and Patterson, 1979; Nelson *et al.*, 2001), implying that off-frequency listening has little effect on threshold in these conditions.

C. Comparison with previous studies

The results of experiment 1 confirm the earlier findings (Harris and Rawnsley, 1953; Harris *et al.*, 1951; Jesteadt

TABLE III. The best-fitting parameter values for the temporal window model for each set of mean data. *c* is the compression exponent, *B* is the breakpoint of the broken-stick nonlinearity, *k* is the detector efficiency, and *A* is the attenuation applied to off-frequency maskers. Values shown in brackets were not varied in the fit. Also shown are the root-mean-squared deviations (the square root of the mean of the squared deviations) of the model predictions from the measured values.

Signal frequency, f_s (Hz)	Masker frequency, f_m (Hz)	Compression exponent, *c*	Break-point, *B* (dB SPL)	Efficiency, *k* (dB)	Attenuation, *A* (dB)	rms error (dB)
250	250	0.24	41.48	2.28	(0)	1.64
250	150	(0.24)	(41.48)	(2.28)	27.44	2.14
500	500	0.25	32.93	2.06	(0)	0.44
500	300	(0.25)	(32.93)	(2.06)	24.87	1.46
4000	4000	0.24	31.34	1.50	(0)	2.47
4000	2400	(0.24)	(31.34)	(1.50)	42.26	3.81

et al., 1982; Lüscher and Zwislocki, 1949; Rawnsley and Harris, 1952): The growth of forward masking with masker level is similar at low and high frequencies. In addition, and in contrast to the results at high signal frequencies (Oxenham and Plack, 1997, 1998, 2000), the growth of masking for off-frequency maskers is not greatly different from that for on-frequency maskers for low signal frequencies. There is less of a *differential* effect of masker frequency for low signal frequencies.

At the 4000-Hz signal frequency, the magnitude of the BM compression estimated here (0.15–0.29 dB/dB in experiment 1, around 0.05–0.37 dB/dB in experiment 2) is similar to that estimated in recent psychophysical studies (Nelson *et al.*, 2001; Oxenham and Plack, 1997) and that measured directly in recent physiological measurements from the basal region of the cochlea (Murugasu and Russell, 1995; Rhode and Recio, 2000; Rhode and Robles, 1974; Robles *et al.*, 1986; Ruggero *et al.*, 1997; Russell and Nilsen, 1997). The results add support to the argument that forward-masking experiments can be used to derive the BM response function in humans.

D. Compression in the apical region of the cochlea

The present results provide evidence for substantial compression at CFs as low as 250 Hz in the human auditory system. It was argued in the Introduction that previous psychophysical measures of low-CF compression depended on the assumption of a linear response to a tone below CF. Consistent with the chinchilla data of Rhode and Cooper (1996), the results of experiment 1 suggest that this assumption is not valid. The shallow growth of masking was present for all masker frequencies relative to the signal frequency, implying that the compression extends over a wide frequency range relative to each CF. Similarly, the TMCs measured in experiment 2 for listeners VD and PP show steep slopes for both on- and off-frequency maskers for the 250-Hz signal.

Two other psychophysical results offer support for the idea that there is considerable compression at low CFs in humans. First, the rate of growth of loudness with level varies little with frequency, except for frequencies below around 100 Hz where the growth becomes steeper (Robinson and Dadson, 1956). Estimates of the function relating physical intensity to perceived magnitude (Buus *et al.*, 1998; Hellman, 1976) are similar to measurements of the BM response to a tone at CF, also showing a more linear response at low levels. In other words, loudness may be a linear function of BM intensity (Schlauch *et al.*, 1998). If this is the case, then the finding of shallow loudness growth with level at low frequencies is suggestive of high cochlear compression at low CFs. Second, Oxenham and Dau (2001) showed recently that there are large effects of the relative phase of harmonics on the amount of masking produced by a complex tone. In a compressive system, the peakier the waveform, the smaller the output rms level compared to the input rms level. An effect on masking of harmonic phase, which alters the envelope of the waveform, is taken as evidence for auditory compression. Oxenham and Dau found large phase effects for signal frequencies as low as 125 Hz.

The question remains, however, as to whether the compression is of cochlear origin, or whether it is the result of more central processes. The frequency dependence of the nonlinearities in forward masking at *high* frequencies strongly points to a source in the cochlea. It is hard to imagine how a single frequency channel could provide a differential response to tones of different frequencies unless the nonlinearities occurred as a consequence of cochlear processing. Once the frequency components combine to determine the response of an auditory nerve fiber with a particular CF, they should then be subject to the same nonlinearities. At low frequencies, however, the compression is not frequency dependent, so it is at least possible that the nonlinearities in the masking data are a consequence of neural processing; for example, saturation in the rate-level functions of auditory neurons. Assuming this is the cause, it seems unlikely that the low-CF saturation is at the level of the auditory nerve. The recordings of Sachs and Abbas (1974) in the cat and Cooper and Yates (1994) in the guinea pig suggest that the rate-level functions of auditory nerve fibers exhibit slightly *less* saturation, if anything, as CF is decreased.

If the nonlinearities in the present psychophysical results are a consequence of cochlear nonlinearities, then it can be said that the compression at CFs of 250 and 500 Hz in the human mechanics appears to be much stronger than in the other mammals (chinchilla, guinea pig, cat) that have been studied so far (Cooper and Yates, 1994; Rhode and Cooper, 1996; Sewell, 1984; Zinn *et al.*, 2000).

V. CONCLUSIONS

(i) The slopes of growth-of-forward-masking functions (signal level versus masker level) are similar for signal frequencies of 250, 500, and 4000 Hz when the masker is at the signal frequency. For masker frequencies below the signal frequency the slopes are steeper at 4000 Hz, but at the two lower frequencies they do not differ greatly from the on-frequency slopes.

(ii) The slopes of temporal masking functions (masker level versus masker–signal interval) are similar for signal frequencies of 250 and 4000 Hz when the masker is at the signal frequency. For masker frequencies below the signal frequency the slopes are shallower at 4000 Hz, but at 250 Hz they do not differ greatly from the on-frequency results.

(iii) These results are consistent with a low-CF compression of around 0.2–0.3 dB/dB that does not vary substantially with stimulation frequency relative to CF. The possibility remains, however, that the compression is not cochlear in origin.

ACKNOWLEDGMENTS

The authors thank David Nelson and an anonymous reviewer for detailed and helpful comments on an earlier version of the manuscript. The research was supported by EPSRC Grant No. GR/N07219.

[1]Some readers may be concerned at the relatively low absolute thresholds for these listeners compared to ANSI (1996) values. Partly, this reflects the higher sensitivity of two of the individuals used in the experiment in comparison to Essex laboratory norms. For 200-ms signals, the absolute thresholds for VD, PP, and EK at 250 Hz were −1.0, 5.3, and 11.7 dB SPL, respectively. This compares to the Essex laboratory norm for 25 listeners at 250 Hz of 8.7 dB SPL. However, even the laboratory norm is considerably less than the ANSI standard of 25.5 dB SPL. The difference may be attributed to the coupler used to estimate sound pressure at the eardrum. In the ANSI calibration a closed NBS-9A coupler was used. This can overestimate the sound pressure at the eardrum by around 10–20 dB at 250 Hz compared to measurements in a Kemar manikin with Zwislocki couplers (Ciechanowski and Cooper, 1976). The Sennheiser HD 580 headphones used in the experiments were calibrated using a Kemar manikin, and were shown to produce an almost flat frequency response between 100 and 1000 Hz. The Essex laboratory norm at 1000 Hz (same 25 listeners) is 0.1 dB SPL, compared to the ANSI standard of 7.0 dB SPL. That result, and the differences in calibration at low frequencies due to specific couplers, are likely to account for differences between the Essex thresholds and ANSI (1996) standards at low frequencies.

[2]Clipping proved a problem at some extreme points. For the 4000-Hz masker, the threshold value for VD at the 70-ms gap was averaged over four estimates and the value for PP at the 55-ms gap was averaged over only two estimates, because five estimates could not be obtained over repeated attempts without the occurrence of clipping. The resulting thresholds are therefore most likely to be underestimates.

ANSI (**1996**). ANSI S3.6-1996, ''Specification for Audiometes'' (American National Standards Institute, New York).

Buus, S., Muesch, H., and Florentine, M. (**1998**). ''On loudness at threshold,'' J. Acoust. Soc. Am. **104**, 399–410.

Ciechanowski, J., and Cooper, W. (**1976**). ''Problems with earphone calibration using a manikin,'' J. Am. Aud. Soc. **2**, 88–94.

Cooper, N. P., and Yates, G. K. (**1994**). ''Nonlinear input-output functions derived from the responses of guinea-pig cochlear nerve fibres: variations with characteristic frequency,'' Hear. Res. **78**, 221–234.

Harris, J. D., and Rawnsley, A. I. (**1953**). ''Patterns of cochlear adaptation at three frequency regions,'' J. Acoust. Soc. Am. **25**, 760–764.

Harris, J. D., Rawnsley, A. I., and Kelsey, P. (**1951**). ''Studies in short-duration auditory fatigue: I. Frequency differences as a function of intensity,'' J. Exp. Psychol. **42**, 430–436.

Hellman, R. (**1976**). ''Growth of loudness at 1000 and 3000 Hz,'' J. Acoust. Soc. Am. **60**, 672–679.

Hicks, M. L., and Bacon, S. P. (**1999**). ''Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing,'' J. Acoust. Soc. Am. **105**, 326–338.

Houtgast, T. (**1972**). ''Psychophysical evidence for lateral inhibition in hearing,'' J. Acoust. Soc. Am. **51**, 1885–1894.

Jesteadt, W., Bacon, S. P., and Lehman, J. R. (**1982**). ''Forward masking as a function of frequency, masker level, and signal delay,'' J. Acoust. Soc. Am. **71**, 950–962.

Johnson-Davies, D., and Patterson, R. D. (**1979**). ''Psychophysical tuning curves: Restricting the listening band to the signal region,'' J. Acoust. Soc. Am. **65**, 765–770.

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

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. (in press).

Lüscher, E., and Zwislocki, J. J. (**1949**). ''Adaptation of the ear to sound stimuli,'' J. Acoust. Soc. Am. **21**, 135–139.

Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Growth of forward masking for sinusoidal and noise maskers as a function of signal delay: Implications for suppression in noise,'' J. Acoust. Soc. Am. **73**, 1249–1259.

Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (**1988**). ''The shape of the ear's temporal window,'' J. Acoust. Soc. Am. **83**, 1102–1116.

Moore, B. C. J., Vickers, D. A., Plack, C. J., and Oxenham, A. J. (**1999**). ''Interrelationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism,'' J. Acoust. Soc. Am. **106**, 2761–2778.

Munson, W. A., and Gardner, M. B. (**1950**). ''Loudness patterns-a new approach,'' J. Acoust. Soc. Am. **22**, 177–190.

Murugasu, E., and Russell, I. J. (**1995**). ''Salicylate ototoxicity: The effects on basilar membrane displacement, cochlear microphonics, and neural responses in the basal turn of the guinea pig cochlea,'' Aud. Neurosci. **1**, 139–150.

Nelson, D. A., Schroder, A. C., and Wojtczak, M. (**2001**). ''A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **110**, 2045–2064.

O'Loughlin, B. J., and Moore, B. C. J. (**1981**). ''Off-frequency listening: effects on psychoacoustical tuning curves obtained in simultaneous and forward masking,'' J. Acoust. Soc. Am. **69**, 1119–1125.

Oxenham, A. J., and Dau, T. (**2001**). ''Towards a measure of auditory-filter phase response,'' J. Acoust. Soc. Am. **110**, 3169–3178.

Oxenham, A. J., and Moore, B. C. J. (**1994**). ''Modeling the additivity of nonsimultaneous masking,'' Hear. Res. **80**, 105–118.

Oxenham, A. J., and Moore, B. C. J. (**1995**). ''Additivity of masking in normally hearing and hearing-impaired subjects,'' J. Acoust. Soc. Am. **98**, 1921–1934.

Oxenham, A. J., and Moore, B. C. J. (**1997**). ''Modeling the effects of peripheral nonlinearity in listeners with normal and impaired hearing,'' in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Hillsdale, NJ), pp. 273–288.

Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.

Oxenham, A. J., and Plack, C. J. (**1998**). ''Suppression and the upward spread of masking,'' J. Acoust. Soc. Am. **104**, 3500–3510.

Oxenham, A. J., and Plack, C. J. (**2000**). ''Effects of masker frequency and duration in forward masking: Further evidence for the influence of peripheral nonlinearity,'' Hear. Res. **150**, 258–266.

Plack, C. J., and Moore, B. C. J. (**1990**). ''Temporal window shape as a function of frequency and level,'' J. Acoust. Soc. Am. **87**, 2178–2187.

Plack, C. J., and Oxenham, A. J. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**, 1598–1608.

Plack, C. J., and Oxenham, A. J. (**2000**). ''Basilar-membrane nonlinearity estimated by pulsation threshold,'' J. Acoust. Soc. Am. **107**, 501–507.

Plack, C. J., Oxenham, A. J., and Drga, V. (**2002**). ''Linear and nonlinear processes in temporal masking,'' Acustica **88**, 348–358.

Rawnsley, A. I., and Harris, J. D. (**1952**). ''Studies in short-duration auditory fatigue: II Recovery time,'' J. Exp. Psychol. **43**, 138–142.
Rhode, W. S., and Cooper, N. P. (**1996**). ''Nonlinear mechanics in the apical turn of the chinchilla cochlea in vivo,'' Aud. Neurosci. **3**, 101–121.
Rhode, W. S., and Recio, A. (**2000**). ''Study of mechanical motions in the basal region of the chinchilla cochlea,'' J. Acoust. Soc. Am. **107**, 3317–3332.
Rhode, W. S., and Robles, L. (**1974**). ''Evidence from Mössbauer experiments for nonlinear vibration in the cochlea,'' J. Acoust. Soc. Am. **55**, 588–596.
Robinson, D. W., and Dadson, R. S. (**1956**). ''A re-determination of the equal-loudness relations for pure tones,'' Br. J. Appl. Phys. **7**, 166–181.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). ''Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and phase responses,'' J. Acoust. Soc. Am. **80**, 1364–1374.
Ruggero, M. A., and Rich, N. C. (**1991**). ''Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane,'' J. Neurosci. **11**, 1057–1067.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Russell, I. J., and Nilsen, K. E. (**1997**). ''The location of the cochlear amplifier: Spatial representation of a single tone on the guinea pig basilar membrane,'' Proc. Natl. Acad. Sci. U.S.A. **94**, 2660–2664.
Sachs, M. B., and Abbas, P. J. (**1974**). ''Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli,'' J. Acoust. Soc. Am. **56**, 1835–1847.
Schlauch, R. S., DiGiovanni, J. J., and Reis, D. T. (**1998**). ''Basilar membrane nonlinearity and loudness,'' J. Acoust. Soc. Am. **103**, 2010–2020.
Sewell, W. F. (**1984**). ''The effects of furosemide on the endocohlear potential and auditory-nerve fiber tuning in cats,'' Hear. Res. **14**, 305–314.
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.
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., Winter, I. M., and Robertson, D. (**1990**). ''Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range,'' Hear. Res. **45**, 203–220.
Zinn, C., Maier, H., Zenner, H.-P., and Gummer, A. W. (**2000**). ''Evidence for active, nonlinear, negative feedback in the vibration response of the apical region of the in-vivo guinea-pig cochlea,'' Hear. Res. **142**, 159–183.

Control of voice fundamental frequency in speaking versus singing

Ulrich Natke,[a] Thomas M. Donath, and Karl Th. Kalveram
Institute of Experimental Psychology, Heinrich-Heine-University, Düsseldorf, Germany

(Received 21 June 2002; accepted for publication 16 December 2002)

In order to investigate control of voice fundamental frequency (F_0) in speaking and singing, 24 adults had to utter the nonsense word ['ta:tatas] repeatedly, while in selected trials their auditory feedback was frequency-shifted by 100 cents downwards. In the speaking condition the target speech rate and prosodic pattern were indicated by a rhythmic sequence made of white noise. In the singing condition the sequence consisted of piano notes, and subjects were instructed to match the pitch of the notes. In both conditions a response in voice F_0 begins with a latency of about 150 ms. As predicted, response magnitude is greater in the singing condition (66 cents) than in the speaking condition (47 cents). Furthermore the singing condition seems to prolong the after-effect which is a continuation of the response in trials after the frequency shift. In the singing condition, response magnitude and the ability to match the target F_0 correlate significantly. Results support the view that in speaking voice F_0 is monitored mainly supra-segmentally and controlled less tightly than in singing. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1543928]

PACS numbers: 43.66.Ba, 43.66.Hg, 43.70.Aj [AL]

I. INTRODUCTION

Auditory control of voice fundamental frequency (F_0) has been investigated with the frequency shift paradigm in many studies before [an overview can be found in Donath *et al.* (2002)]. From the very beginning, continuous vocalization has been studied (e.g., Elman, 1981; Burnett *et al.*, 1997; Larson, 1998). In this approach, subjects usually produced a vowel for a duration of about 5 s. After a randomly determined period, auditory feedback provided via headphones was shifted in frequency. Typically subjects reacted with a change in voice F_0 in the opposite direction of the frequency shift (''opposing response''), which indicates a closed-loop negative feedback system compensating deviations between intended and perceived pitch. However, inter-individual differences in responses were large and responses in the direction of the frequency shift (''following responses'') were also observed. There is some evidence that opposing responses occur with a short latency of 100–150 ms, whereas following responses have a longer latency of 250–600 ms (Larson, 1998). The presence of two responses was confirmed by Hain and colleagues (2000), who found that the direction of the second response, but not of the first response, can be modified by instruction. The first response therefore indicates a negative feedback system stabilizing voice F_0 automatically. The second response may reflect a voluntary mechanism, which adjusts voice F_0 to match an (supposed) external reference.

In order to investigate voice F_0 control in speaking, which is characterized by rapidly repeated onsets and offsets of phonation, Natke and Kalveram (2001) and Donath *et al.* (2002) utilized a perturbation paradigm which was used previously in investigating the control of vowel duration with delayed auditory feedback (e.g., Kalveram and Jäncke, 1989; Jäncke, 1991; Natke, 1999). In this paradigm, subjects utter a nonsense word repeatedly and auditory feedback is modified in randomly selected trials. These trials are compared to preceding trials in which auditory feedback is not altered. The prosodic pattern of the nonsense word is varied, so that effects of feedback manipulation on short unstressed and long stressed syllables in different positions within the word can be investigated.

In the studies of Natke and Kalveram (2001) and Donath *et al.* (2002) only opposing responses occurred with a latency of about 160 ms due to frequency-shifted auditory feedback. This indicates that in speaking only the first, involuntary responses occur. Because of the short latency, voice F_0 cannot be controlled auditorily within short syllables and responses in long stressed syllables come into effect late. The response takes place over successive syllables, even when phonation stops between them. Furthermore, an after-effect in trials after termination of frequency shift is found, indicating that the response persists for several seconds. Therefore the system is able to adjust voice F_0 in later syllables by monitoring current syllables. The auditory-vocal system controlling voice F_0 therefore seems to operate mainly at the supra-segmental level, not at the syllabic level.

In speaking, response magnitudes of about 50 cents were found, although a complete compensation would have required a change of 100 cents in voice F_0. These magnitudes correspond to magnitudes found in continuous vocalization (mostly 30–60 cents using frequency shifts of 100–600 cents). Thus there seems to be a limiting property of the audio-vocal system, which prevents responses from exceeding a certain magnitude. Natke and Kalveram (2001) and Donath *et al.* (2002) explain these findings by pointing out that for the comprehension of speech sound duration, formants and formant transitions seem to be more important than voice F_0, at least in languages like English or German. Control of voice F_0 at a supra-segmental level but not of

[a]Electronic mail: natke@uni-duesseldorf.de

absolute pitch within syllables is important to encode nonverbal information. Therefore, a tight control of voice F_0 is unnecessary in speaking.

However, in singing, tight control of voice F_0 seems preferable. Whereas in speaking continuous changes of voice F_0 occur, singing is usually characterized by matching absolute values of pitch mainly stepwise, corresponding to musical notes (apart from exceptions such as performing a glissando, singing with vibrato or singing blue notes making up the blues scale, where pitch is not fixed precisely but varies). In speaking an external reference for voice F_0 does not exist, but in singing, for instance, in a choir or accompanied by musical instruments, an external reference is provided. A deviation between the F_0 of the own voice and the external reference must be compensated. This is easily achieved by trained singers; for example, trained singers can match a reference tone of 440 Hz with an accuracy of less than 1 Hz (Sundberg, 1987). Also, solo singing requires some form of a reference, so that individual notes of scales can be sung with the correct pitch. Even untrained singers are usually aware of when they do not produce the right pitch, resulting in the characteristic ''bad'' glissando while trying to match their voice F_0 with the target F_0. This reference must be represented internally. In speaking such an internal reference for voice F_0 must be used, too, because in the perturbation studies using test words compensating responses were found. However, in singing this internal reference obviously is more prominent. Consequently in singing, a frequency shift of auditory feedback should be compensated almost completely. Evidence for that is reported by Burnett *et al.* (1997) who had trained singers sing musical scales. In some subjects a complete compensation of unanticipated frequency shifts of 100 cents was observed. However, only exemplary data is reported.

In summary, it is suggested that in speaking an internal reference for pitch plays a subordinate role for controlling voice F_0 leading to the limiting property, whereas the audiovocal system is able to operate very effectively in controlling voice F_0, if a precise reference is provided as in singing. Therefore the response to frequency-shifted auditory feedback should be greater in singing than in speaking. The present study was designed to test this hypothesis.

II. METHOD

A. Subjects

Twenty-four adult German native speakers (9 women, 15 men) between 21 and 33 years of age participated in this study ($M = 25.9$ years, s.d.$= 3.92$ years). All subjects passed a hearing screening at 20 dB HL (Hortmann DA 323, Neckartenzlingen, Germany). No subject reported a current speech or language disorder and none was trained as a professional singer.

B. Apparatus

Voice frequency was shifted using a commercial device (DFS 404, Casa Futura Technologies, Boulder, Colorado), which works on a digital basis (sampling frequency: 32 kHz, sampling resolution: 14 bits). The device had been modified by installing a relay to enable remote-switching between nonaltered auditory feedback and frequency-shifted auditory feedback. Auditory feedback was provided through sealed headphones of a headset (Blackhawk, DSP 5DX, Flightcom, Portland, Oregon), which attenuate air-conducted sound by approximately 24 dB SPL. Feedback volume was calibrated so that a sinusoidal tone of 440 Hz with 75 dB (A) at the microphone led to a feedback volume of 70 dB (A) in the headphones. Subjects perceived this volume as a normal feedback volume. In order to mask bone conduction, during the whole experiment low-pass-filtered white noise ($f_c = 900$ Hz) was binaurally added at an intensity of 65 dB (A). Since there is no simple approach to physical measurement of masking efficiency, our own and the subjects' observations served as an indicator. When the auditory feedback was turned off suddenly while the masking noise continued, subjects reported that to them it seemed as if they had stopped speaking, respectively singing, because they could no longer hear their voice. This indicates that the masking level was sufficient to mask any speech sounds not presented through headphones. However, the masking level was still low enough for subjects to hear their voice clearly and distinctly over the noise.

In order to determine voice F_0, the vibrations of the vocal folds were recorded directly with an electroglottograph (EGG; Laryngograph, Kay Elemetrics, Pine Brook, New Jersey). Control of the experiment and data acquisition was automatized by a commercial personal computer with a soundcard. The computer switched the frequency shift device between nonaltered and frequency-shifted auditory feedback and digitized the EGG signal with a sampling rate of 11 025 Hz and a sampling resolution of 16 bits.

C. Procedure

Subjects had to utter the nonsense word ['ta:tatas] with a speech rate indicated by a rhythmic sequence presented via headphones before speaking. The first long-stressed syllable was represented by a sound with a duration of 400 ms. Each of the following two sounds represented the unstressed syllables and were 200 ms long. Sounds were separated by pauses of 40 ms. The sequence was presented twice; afterwards 3.5 s of silence allowed the subject to speak the word. The sequence was presented automatically while the subject sat alone in a sound-isolated chamber.

Two conditions were realized which differed regarding the sound sequence and the instruction. In the ''speaking condition,'' the sequence consisted of white noise. Therefore, no kind of reference pitch for voice F_0 was provided. Subjects were asked to speak clearly and with normal volume, but no specific instruction regarding the fundamental frequency was given. In the ''singing condition,'' the sequence consisted of piano notes. Subjects were instructed to match the pitch of the piano notes while singing the nonsense word. However, the piano notes were presented before subjects' utterance, not during it. The fundamental frequency of the piano notes was varied based on the sex of the subjects: females: 233 Hz, males: 123 Hz. These frequencies were chosen because they are very close to the average voice F_0

of women and men and resemble the musical notes a#3 and b2. In neither condition were subjects informed that the feedback would be modified from time to time.

In a training phase preceding each of the two phases of the experiment, subjects had to utter or sing the nonsense word simultaneously, while the target white noise or target piano notes, respectively, were presented in a loop. This was done until they produced the word correctly at least five times in succession, as judged by the experimenter. Generally, subjects reached this criterion within ten trials.

The experimental procedure consisted of 30 trials for the speaking condition and 30 trials for the singing condition. In 20% of the trials frequency was shifted downwards by 100 cents (one semitone), resulting in six frequency-shifted trials per condition. The frequency shift was activated at the beginning of the rhythmic sequence, so that auditory feedback, but not the sequence sounds themselves, were frequency-shifted from the very beginning of subjects' utterances. Subjects then produced the entire word with frequency-shifted auditory feedback. Frequency shift was deactivated before the next trial began. Trials with frequency shift were randomly selected with the limitations that at least two trials with nonaltered auditory feedback had to precede a trial with frequency shift and that the last trial of an experimental phase was always conducted with nonaltered auditory feedback. Therefore it was guaranteed that before each trial with frequency shift there was always a trial with nonaltered feedback, which did not immediately follow frequency shift. In this way trials immediately before frequency shift could be used as the reference to calculate deviations in voice F_0, while trials immediately after frequency shift could be used to investigate the after-effect. The sequence of the two conditions was randomized as well.

D. Data analysis

Data analysis was almost identical to Donath *et al.* (2002), in which the method is described in more detail. Based on the EGG-signal, momentary voice F_0 and vowel duration were calculated. Contours of F_0 were obtained for vowels of the first two syllables of the nonsense word, the first being long stressed and the second being unstressed. First F_0 contours with a fixed resolution of 0.1 ms were determined. This was achieved by linear interpolation of the momentary frequency of each vocal fold period. In some cases there were irregular high-frequency fluctuations in voice F_0 contours due to problems with analyzing EGG-signals with a poor signal-to-noise ratio. Therefore voice F_0 contours were smoothed with a moving average over 10 ms. Phonation onset was defined as the first closing instant detected in the EGG-signal. Mean voice F_0 contours for individual subjects were truncated at the end according to the shortest vowel duration produced by this subject. Thus, all portions of the contour were based on the total number of trials. To avoid F_0 contours being truncated excessively based on unusually short vowels, trials in which vowel duration was 25% shorter compared to all other trials were discarded. A total of 30 trials (speaking condition: 18, singing condition: 12) were discarded because of insufficient length or artifacts, resulting in a total of 834 valid trials. Despite the training, one subject produced the second syllable with an unusual vowel duration shorter than 20 ms during the actual experiment, so the data of this subject were excluded from analysis of the second syllable.

Mean voice F_0 contours of vowels in words before (PRE), during (FAF), and after (POST) trials with frequency-shifted auditory feedback were calculated. Remaining trials were discarded. PRE-trials were used as reference trials, to which FAF- and POST-trials were compared for each subject. In order to eliminate interindividual variance due to the subjects' habitual voice F_0 (especially due to the sex of the subjects), for each data point the difference in cents was calculated between F_0 contours in PRE- and FAF-trials as well as in PRE- and POST-trials. Resulting contours therefore reflect the deviation from normal production of voice F_0 during frequency shift and after its termination. Finally, individual subjects' contours were averaged to obtain the group data.

Based on previous findings (see Donath *et al.*, 2002) it is hypothesized that voice F_0 is higher in the final portion of the first long stressed syllable and the initial and final portion of the second unstressed syllable in FAF-trials. Due to the after-effect voice F_0 should also be higher in the initial portion of the first syllable in POST-trials. The initial portion of syllables was defined as the interval 25–50 ms after vowel onset and the final portion was defined as the interval 200–225 ms in the first long stressed syllable and 75–100 ms in the second short unstressed syllable. These intervals were chosen to be as late as possible and yet featuring voice F_0 data for most subjects. Means for these intervals were calculated for each subject and one-tailed Wilcoxon-signed-rank-tests were calculated to test for responses. A significance level of $\alpha = 5\%$ was chosen and corrected according to Bonferroni to $\alpha' = \alpha/8 = 0.006\,25$. For intervals for which no differences were assumed, additional p-values were calculated two-tailed and interpreted as measures of effect. Response latencies for syllables, in which changes in voice F_0 were indicated by differences between the initial and final portion, were determined using the Castellan change-point test (Siegel and Castellan, 1988).

Furthermore, it is hypothesized that the response in the singing condition is greater than in the speaking condition. The difference between the speaking and singing condition was tested by comparing the response magnitudes in the interval 25–75 ms of the second syllable because in this interval the response is at its maximum and relatively stable as found by Donath *et al.* (2002).

The same interval was chosen in order to investigate whether a relationship exists between the ability to accurately match a note which is provided externally and the magnitude of compensation under frequency shift. For each subject the response magnitude was calculated as the deviation of voice F_0 contours of FAF-trials from PRE-trials in the interval 25–75 ms of the second syllable in the speaking as well as in the singing condition. The same interval in PRE-trials in the singing condition was chosen to calculate the deviation of voice F_0 from the fundamental frequency of the piano notes (target F_0), which were presented before the utterance to provide the target F_0. The absolute value of this

TABLE I. Mean vowel duration in speaking versus singing of the first (=long stressed) syllable and the second (=unstressed) syllable of the nonsense word ['ta:tatas] (standard deviation in brackets). Vowel duration was averaged over all PRE-, FAF- and POST-trials.

	Vowel duration (s.d.) (ms)	
Condition	1st syllable (long stressed)	2nd syllable (unstressed)
Speaking	249.5 (33.5)	89.2 (22.9)
Singing	278.4 (36.5)	93.4 (27.5)

deviation indicates the accuracy of note matching. Spearman's rho correlation was calculated for the response magnitude and the absolute deviation from the target F_0.

III. RESULTS

A. Response latency and response magnitude during frequency shift

Table I reports the vowel duration of the first (=long stressed) syllable and the second (=unstressed) syllable of the nonsense word ['ta:tatas] averaged over PRE-, FAF- and POST-trials. The vowel duration of the stressed syllable was nearly three times longer than the vowel duration of the unstressed syllable. Whereas in the unstressed syllable the vowel duration in the speaking and in the singing condition do not differ ($p_{2\text{-tailed}}=0.280$, $Z=-1.080$, $n=24$), the vowel duration of the long stressed syllable is approximately 29 ms longer in the singing than in the speaking condition ($p_{2\text{-tailed}}=0.002$, $Z=-3.148$, $n=24$). Therefore the prolongation of vowels in singing, which sometimes is obvious, was found in the present experiment, however restricted to long stressed syllables.

Figure 1 shows the mean voice F_0 contours of the first syllable in PRE-trials. In all graphs, 0 ms corresponds to the first detected glottal closing, which defines onset of phonation. Although piano notes with a fundamental frequency close to the average voice F_0 of women and men were used, subjects had a higher voice F_0 in the singing condition compared to the speaking condition. The mean difference is 11 Hz corresponding to 119 cents. Until 75 ms after vowel onset voice F_0 in the speaking condition decreases stronger than in the singing condition. In the remaining part there are no systematic differences apart from the offset. Therefore it seems appropriate to normalize the data to the cent-scale in order to make comparisons between subjects with different habitual voice F_0 (especially between women and men) possible.

Figure 2 shows the deviation of voice F_0 contours of FAF-trials from PRE-trials. In the interval 25–50 ms after vowel onset [Fig. 2(a)] mean voice F_0 of the first syllable does not differ between PRE- and FAF-trials, neither in the speaking ($p_{2\text{-tailed}}=0.797$, $Z=-0.257$, $n=23$) nor in the singing condition ($p_{2\text{-tailed}}=0.189$, $Z=-1.314$, $n=23$). In the speaking condition the response begins after 154.6 ms, in the singing condition after 142.1 ms, as calculated with the Castellan change-point test. In the interval 200–225 ms after vowel onset, mean voice F_0 of the first syllable is 17.4 cents higher in FAF-trials in the speaking condition (s.d. $=34.0$ cents; $p_{1\text{-tailed}}=0.005$, $Z=-2.571$, $n=24$), while in the singing condition voice F_0 is 35.5 cents higher than in PRE-trials (s.d.$=28.8$ cents; $p_{1\text{-tailed}}<0.001$, $Z=-3.954$, $n=23$).

In both conditions, mean voice F_0 of the second syllable is also higher in FAF-trials compared to PRE-trials [Fig. 2(b)]. In the speaking condition the difference in the interval 25–50 ms after vowel onset is 45.4 cents (s.d. $=37.9$ cents; $p_{1\text{-tailed}}<0.001$, $Z=-3.832$, $n=23$), in the singing condition 64.6 cents (s.d.$=31.1$ cents; $p_{1\text{-tailed}}<0.001$, $Z=-4.106$, $n=23$). In the interval 75–100 ms after vowel onset, mean voice F_0 of the second syllable is 50.2 cents higher in FAF-trials in the speaking condition (s.d. $=38.8$ cents; $p_{1\text{-tailed}}<0.001$, $Z=-3.582$, $n=19$), while in the singing condition voice F_0 is 64.2 cents higher than in PRE-trials (s.d.$=27.0$ cents; $p_{1\text{-tailed}}<0.001$, $Z=-3.883$, $n=20$).

In the interval 25–75 ms of the second syllable, which was chosen for comparison of the speaking and singing condition, response magnitude in the singing condition ($M=66.1$ cents, s.d.$=30.4$ cents) is higher than in the speaking condition ($M=46.7$ cents, s.d.$=37.3$ cents) ($p_{1\text{-tailed}}=0.013$, $Z=-2.220$, $n=23$).

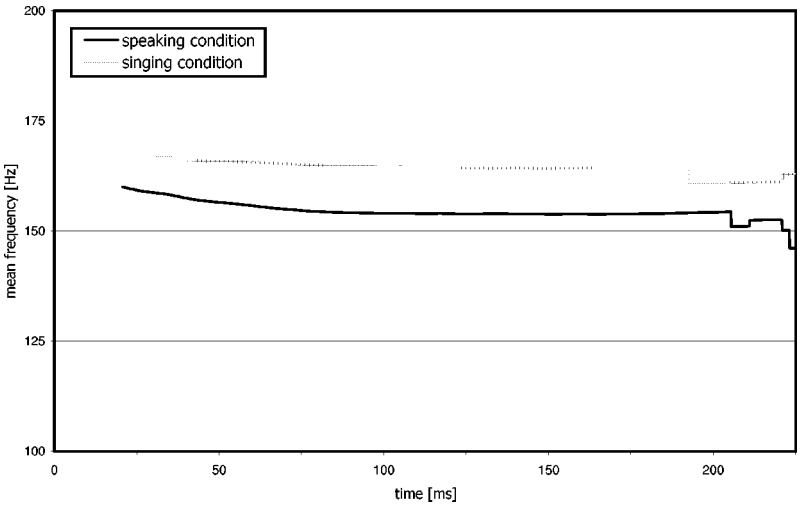


FIG. 1. Mean voice F_0 contours of the first syllable (=long stressed) of the nonsense word ['ta:tatas] in trials before frequency shift (PRE). 0 ms is the onset of phonation, as defined by the first glottal closing. The contour begins at the point for which voice F_0 data were available for all subjects. Steps at the end of these averaged graphs are the result of single subjects stopping their vocalization.

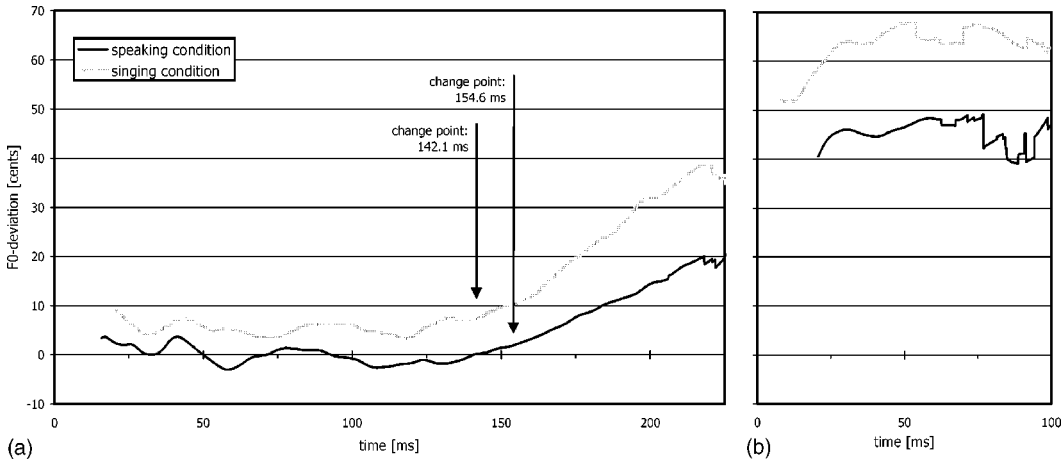


FIG. 2. Average deviation of voice F_0 contours of the first syllable [=long stressed (a)] and the second syllable [=unstressed (b)] of the nonsense word ['ta:tatas] during frequency shift (FAF) from trials before frequency shift (PRE) in cents. Black line speaking condition, gray line singing condition. See caption for Fig. 1 for additional explanation.

B. Response latency and response magnitude after termination of frequency shift

In Fig. 3 the voice F_0 deviations of POST-trials from PRE-trials are shown. In the interval 25–50 ms after vowel onset [Fig. 3(a)], p-values do not reach the Bonferroni corrected significance level, but indicate the tendency that mean voice F_0 of the first syllable is 10.9 cents higher in POST-trials than in PRE-trials in the speaking condition (s.d. $=21.2$ cents; $p_{1\text{-tailed}}=0.014$, $Z=-2.200$, $n=24$) and 16.8 cents higher in the singing condition (s.d. $=33.1$ cents; $p_{1\text{-tailed}}=0.016$, $Z=-2.143$, $n=24$). In the interval 200–225 ms after vowel onset mean voice F_0 does not differ between POST- and PRE-trials in the speaking condition ($p_{2\text{-tailed}}=0.145$, $Z=-1.457$, $n=24$). The response ends after 169.5 ms. However, in the singing condition, voice F_0 is 15.5 cents higher (s.d.$=18.5$ cents; $p_{2\text{-tailed}}<0.001$, $Z=-3.254$, $n=23$).

In the speaking condition, mean voice F_0 of the second syllable [Fig. 3(b)] does not differ between POST-and PRE-trials in the interval 25–50 ms after vowel onset ($p_{2\text{-tailed}}=0.465$, $Z=-0.730$, $n=23$). In the singing condition voice F_0 is 8.2 cents higher (s.d.$=16.3$; $p_{2\text{-tailed}}=0.007$, $Z=-2.677$, $n=23$). In the interval 75–100 ms after vowel onset, mean voice F_0 does not differ between POST- and PRE-trials, neither in the speaking condition ($p_{2\text{-tailed}}=0.629$, $Z=-0.483$, $n=19$) nor in the singing condition ($p_{2\text{-tailed}}=0.279$, $Z=-1.083$, $n=20$). The response ends after 57.0 ms in the singing condition.

C. Relationship between accuracy of note matching and response magnitude

In order to investigate whether a relationship exists between the ability to match a note accurately and the magnitude of compensation under frequency shift, absolute deviation from target F_0 and response magnitude as defined in Sec. II D was calculated for each subject. Four subjects deviate from the fundamental frequency of the piano notes more than one semi-tone in PRE-trials. With deviations of 190, 287, 440, and 526 cents these subjects show virtually no ability to match an external pitch with their voice. These exceptionally large deviations pose a problem as far as the calculation of correlation is concerned because of the large influence of such outliers in a small sample. Consequently,

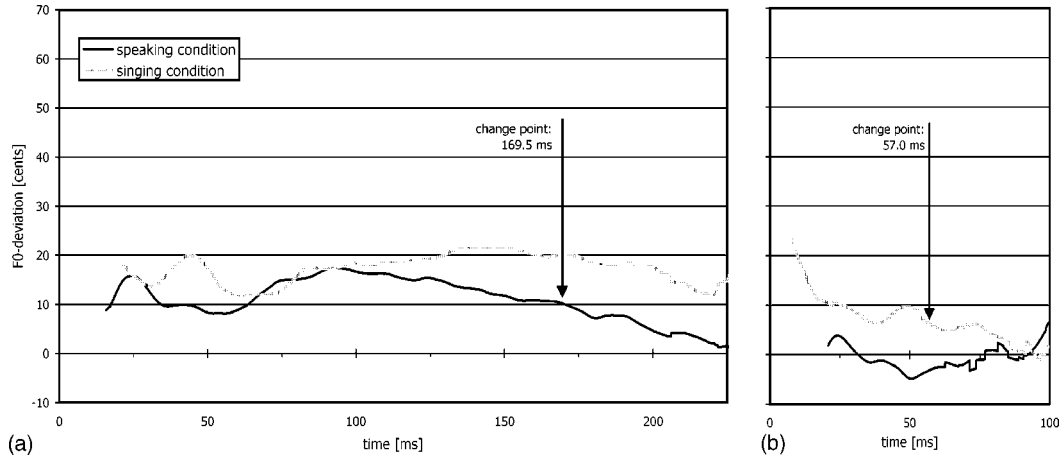


FIG. 3. Average deviation of voice F_0 contours of the first syllable [=long stressed, (a)] and the second syllable [=unstressed, (b)] of the nonsense word ['ta:tatas] after termination of frequency shift (POST) from trials before frequency shift (PRE) in cents. Black line speaking condition, gray line singing condition. See caption for Fig. 1 for additional explanation.

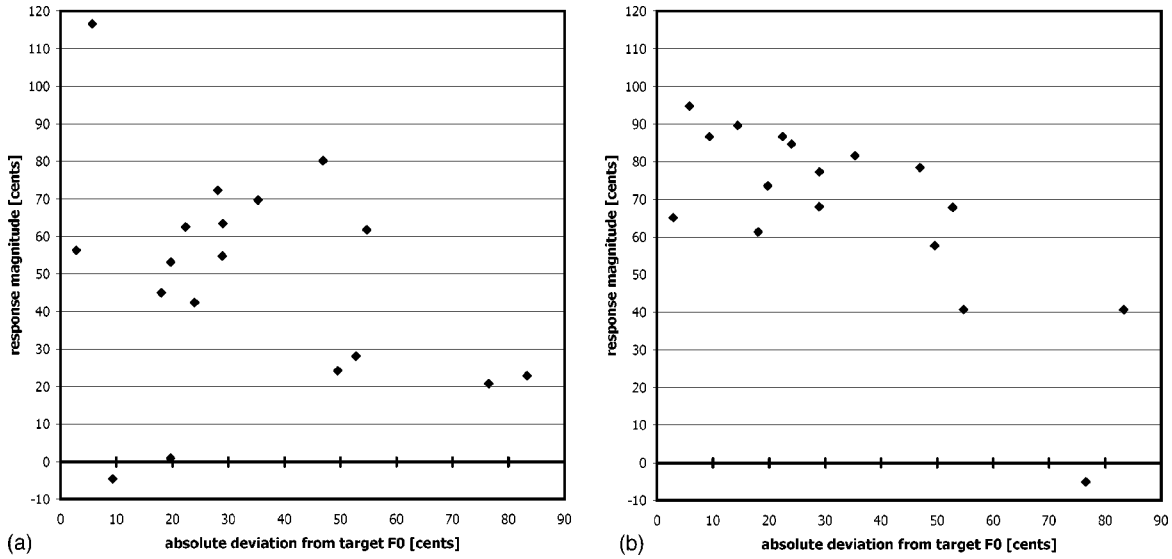


FIG. 4. Individual absolute deviation from the target F_0 (an indication of the accuracy of note matching) and response magnitude in the speaking (a) and in the singing condition (b). Response magnitude was calculated as the mean difference between FAF- and PRE-trials over the interval 25–75 ms after onset of phonation in the second short unstressed syllable. Absolute deviation of target F_0 was calculated in the same interval as the absolute deviation of voice F_0 in PRE-trials from the fundamental frequency of the piano notes presented as the target F_0 in the singing condition.

these four subjects were excluded from the correlation analysis, even though they do show responses to frequency shift. Therefore it should be noted that the results reported here should only be generalized to individuals who show at least some ability to match external notes. Response magnitude and absolute deviation from target F_0 are plotted against each other in Fig. 4. Whereas in the speaking condition [Fig. 4(a)] no significant correlation was found (Spearman's rho $=-0.004$, $p_{1\text{-tailed}}=0.494$, $n=19$), in the singing condition [Fig. 4(b)] response magnitude and absolute deviation from the target F_0 correlate inversely (Spearman's rho $=-0.469$, $p_{1\text{-tailed}}<0.022$, $n=19$). Therefore, in singing, subjects who show better note matching abilities also show greater compensation during frequency shift. However, correlational analysis does not indicate a confoundation of the ability to match the target note and the actual voice F_0 (Spearman's rho $=-0.263$, $p_{2\text{-tailed}}<0.276$, $n=19$).

IV. DISCUSSION

Regarding voice F_0 control in speaking, the results of the present study confirm those of Donath *et al.* (2002). Auditory feedback is used to control voice F_0 in vowel production by a negative-feedback system. Because of the long response latency of approximately 150 ms, deviations between intended and perceived voice F_0 cannot be compensated during the first half of long stressed syllables. Unstressed syllables are therefore too short for a compensation to occur. In subsequent syllables the response persists and reaches its maximum. Therefore control of voice F_0 is not interrupted by the onset and offset of phonation between syllables of a single word. Adjustments of voice F_0 even persist when speaking is paused for several seconds. Therefore, auditory feedback is not used to regulate voice F_0 within syllables, rather it is used for the adjustment of voice F_0 suprasegmentally in terms of prosody.

Except for response magnitude and latency of aftereffect, results in the singing condition are comparable to those of the speaking condition. Thus there is evidence that the same negative feedback system which is used for automatic correction of voice F_0 during speaking is used as well during singing. In the speaking condition, the frequency shift was compensated to an extent of 47% of the frequency shift, demonstrating the limiting property of the auditory-vocal system shown in previous studies. However, in the singing condition the compensation was significantly higher, reaching 66% of the frequency shift. One might object that in this experiment voice F_0 was higher in the singing condition than in the speaking condition, so therefore the frequency shift of 100 cents resulted in a larger absolute shift (measured in Hz) in the singing condition. However, the relative frequency shifts (in cents) were exactly the same and therefore the subjects' (relative) perception in both conditions, too. Furthermore, Burnett *et al.* (1998) found no relationship between frequency shift magnitude and response magnitude. This indicates that the difference in response magnitude between singing and speaking found in this study is rather a result of the conditions, not of differences in frequency shift magnitude in Hz.

The piano notes were presented as the target F_0 just before the utterance, but not during it. Therefore there remained only an internal reference for voice F_0. Results support the hypothesis that forcing an internal reference contributes to the auditory-vocal system being effective in regulating voice F_0. A future study might have subjects sing at a self-selected pitch and omit an external reference. This way it could be investigated whether the instruction to sing is sufficient for forcing an internal reference.

Even in the singing condition, no total compensation of the frequency shift was realized. The ability to match the target F_0 provided by the piano notes varied greatly. The ability to match the target F_0 correlates significantly with the

response magnitude in the singing condition. This means that subjects who have better control over their voice and/or better auditory perception, thus enabling them to match external target frequencies more accurately, also compensate perceived deviations of their voice F_0 better in singing. Therefore the mean response magnitude of 66% of the frequency shift in the singing condition might reflect the distribution of the ability of note matching in the sample. A future study comes to mind easily: When singers and nonsingers are compared regarding their response magnitude during frequency-shifted auditory feedback, singers might show a better compensation in case they match the target notes more accurately than nonsingers. Furthermore, response latency of singers might be shorter than that of nonsingers. This seems possible because musicians have superior preattentive auditory processing abilities (Koelsch *et al.*, 1999). Finally it would be interesting to investigate whether an improvement of the ability of note matching leads to greater responses. Such studies would answer the question whether training could influence the system regulating voice F_0, which seems to work involuntarily and reflexlike (Larson, 1998).

There seems to be no relationship between the ability to match the target F_0 and the response magnitude in the speaking condition. Therefore, the ability for precise voice F_0 control principally exists (as shown by the correlation between the ability to match the target F_0 and the response magnitude in the singing condition), but is not used effectively in speaking. As stated in the Introduction, it is assumed that voice F_0 is not monitored very tightly in speaking because it is not necessary for comprehension of speech, at least in nontone languages. Future studies might address differences between tone and nontone languages in responses to frequency-shifted auditory feedback.

Response latencies in frequency shift trials were comparable in the speaking and the singing condition and are about 150 ms. In trials after termination of frequency shift, voice F_0 seems to be increased from the very beginning of phonation. Voice F_0 therefore is controlled continuously beyond pauses of phonation within words as well as between words which are separated by several seconds. The lasting effect after returning to nonaltered auditory feedback indicates that voice F_0 adapts in a feedforward fashion on a supra-segmental level. In the speaking condition, the after-effect comes to an end within the first syllable after about 170 ms as found by Donath *et al.* (2002) previously. In a continuous vocalization task nearly the same latencies for onset and offset of response as in the present study were found (Larson *et al.*, 2001). Whereas the onset of response is characterized by a relatively fast increase of voice F_0, the decrease of voice F_0 seems to take place more slowly. In the singing condition the after-effect continues to the second syllable. The change point in the second syllable is 57 ms, so the total latency of the after-effect in the singing condition adds to 415 ms with the vowel duration of the first syllable of about 278 ms and the estimated duration of the second consonant of 80 ms. In singing the compensation starts as quickly as in speaking, but remains longer, even when there is no deviation between intended and perceived voice F_0 anymore. In singing, of course, it is important to compensate deviations quickly. Therefore it seems somewhat surprising that the after-effect persists longer, i.e., the latency of correction is higher. Some singers report that they ''glide'' into notes right up from below. Maybe the other direction from top to bottom is less natural and therefore causes the higher latency of the after-effect.

V. CONCLUSION

The study demonstrates how the paradigm of frequency-shifted auditory feedback during speaking versus singing of nonsense words can be applied to get insight into segmental and supra-segmental processes controlling voice F_0. Results show that in speaking voice F_0 is controlled to a lesser extend than in singing and that the degree of note matching ability influences the degree of compensation to frequency shift in singing.

Burnett, T. A., Freedland, M. B., and Larson, C. R. (**1998**). ''Voice F0 responses to manipulations in pitch feedback.'' J. Acoust. Soc. Am. **103**, 3153–3161.

Burnett, T. A., Senner, J. E., and Larson, C. R. (**1997**). ''Voice F0 Responses to pitch-shifted auditory feedback: A preliminary study,'' J. Voice **11**, 202–211.

Donath, T. M., Natke, U., and Kalveram, K. T. (**2002**). ''Effects of frequency-shifted auditory feedback on voice F_0 contours in syllables,'' J. Acoust. Soc. Am. **111**, 357–366.

Elman, J. L. (**1981**). ''Effects of frequency-shifted feedback on the pitch of vocal production,'' J. Acoust. Soc. Am. **70**, 45–50.

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., and Kenney, M. K. (**2000**). ''Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex.'' Exp. Brain Res. **130**, 133–141.

Jäncke, L. (**1991**). ''The 'audio-phonatoric coupling' in stuttering and non-stuttering adults: Experimental contributions,'' in *Speech Motor Control and Stuttering*, edited by H. F. M. Peters, W. Hulstijn, and C. W. Starkweather (Elsevier Scientific, Amsterdam), pp. 171–180.

Kalveram, K. T., and Jäncke, L. (**1989**). ''Vowel duration and voice onset time for stressed and nonstressed syllables in stutterers under delayed auditory feedback condition,'' Folia Phoniatr. **41**, 30–42.

Koelsch, S., Schröger, E., and Tervaniemi, M. (**1999**). ''Superior attentive and pre-attentive auditory processing in musicians,'' NeuroReport **10**, 1309–1313.

Larson, C. R. (**1998**). ''Cross-modality influences in speech motor control: The use of pitch-shifting for the study of *F*0 control,'' J. Commun. Disorders **31**, 489–503.

Larson, C. R., Burnett, M. L., Bauer, J. J., Kiran, S., and Hain, T. C. (**2001**). ''Comparison of voice F_0 responses to pitch-shift onset and offset conditions,'' J. Acoust. Soc. Am. **110**, 2845–2848.

Natke, U. (**1999**). ''Die Kontrolle der Phonationsdauer bei stotternden und nichtstotternden Personen: Einfluß der Rückmeldelautstärke und Adaptation (Control of phonation in stuttering and nonstuttering persons: Influence of feedback loudness and adaptation),'' Sprache-Stimme-Gehör **23**, 198–205.

Natke, U., and Kalveram, K. T. (**2001**). ''Fundamental frequency under frequency shifted auditory feedback of long stressed and unstressed syllables,'' J. Speech Lang. Hear. Res. **44**, 577–584.

Siegel, S., and Castellan, N. J. (**1988**). *Nonparametric Statistics for the Behavioral Sciences*, 2nd ed. (McGraw-Hill, Boston).

Sundberg, J. (**1987**). *The Science of the Singing Voice* (Northern Illinois U.P., Dekalb, IL).

Informational masking caused by contralateral stimulation[a)]

Gerald Kidd, Jr.,[b)] Christine R. Mason, and Tanya L. Arbogast
Hearing Research Center, Sargent College, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02132

Douglas S. Brungart
Air Force Research Laboratory, AFRL/HECB, Wright-Patterson Air Force Base, Ohio 45433

Brian D. Simpson
Veridian, 5200 Springfield Pike, Suite 200, Dayton, Ohio 45431

(Received 28 October 2002; revised 11 December 2002; accepted 16 December 2002)

Although informational masking is thought to reflect central mechanisms, the effects are generally much stronger when the target and masker are presented to the same ear than when they are presented to different ears. However, the results of a recent study by Brungart and Simpson [J. Acoust. Soc. Am. **112**, 2985–2995 (2002)] indicated that a speech masker that is presented contralateral to a speech signal can produce substantial amounts of informational masking when a second speech masker is played simultaneously in the same ear as the signal. In this study, we conducted a series of experiments that paralleled those of Brungart and Simpson but used a pure-tone signal and multitone informational maskers in a detection task. Both the signal and the maskers were played as sequences of short bursts in each observation interval. The maskers were arranged in two types of spectrotemporal patterns. One type of pattern, called ''multiple-bursts same'' (MBS), has previously been shown to produce very large amounts of informational masking while the other type of pattern, called ''multiple-bursts different'' (MBD), has been shown to produce very small amounts of informational masking. Several conditions of ipsilateral, contralateral, and combined presentation of these maskers were tested. The results showed that presentation of the MBS masker in the contralateral ear produced a substantial amount of informational masking when the MBD masker was simultaneously presented to the ipsilateral ear. The results supported the earlier findings of Brungart and Simpson indicating that listeners are unable to selectively focus their attention on a single ear in some complex dichotic listening conditions. These results suggest that this contralateral masking effect is not restricted to speech and may reflect more general limitations on processing capacity. Further, it was concluded that the magnitude of the contralateral masking effect was related both to the informational masking value of the contralateral masker *and* the complexity of the stimulus and/or task in the ear in which the signal was presented. [DOI: 10.1121/1.1547440]

PACS numbers: 43.66.Dc, 43.66.Lj, 43.66.Rq, 43.66.Pn [LRB]

I. INTRODUCTION

In a recent experiment, Brungart and Simpson (2002) measured speech recognition for a monaurally presented speech signal in the presence of one or two speech maskers. When a single masker was presented in the opposite ear from the signal (contralateral masker), the intelligibility of the signal was near 100%. That result was consistent with the findings of the classic dichotic listening experiments of Cherry (1953), Broadbent (1958), and others (for recent reviews, see Yost, 1997 and Bronkhorst, 2000). When a single masker was presented in the same ear as the signal (ipsilateral masker), speech recognition scores ranged from near 100% correct at a signal-to-masker ratio (S/M) of 8 dB to about 70% correct at a S/M of −8 dB. That finding was also expected from previous work on speech-on-speech masking (e.g., Carhart *et al.*, 1959; Festen and Plomp, 1990). However, when one masker was presented ipsilateral to the signal and a second unrelated masker was played contralateral to the signal, additional decrements as large as 40 percentage points (*re*: the ipsilateral-only masker) in the recognition of the signal occurred. To our knowledge, this was a new and unexpected finding. Brungart and Simpson interpreted this apparent inability of listeners to ignore an irrelevant contralateral masker as reflecting limitations on the attentional resources available for segregating speech sounds. Listeners were able to segregate the signal from a single masker played in the contralateral ear almost perfectly, and could segregate the signal from a single masker in the ipsilateral ear reasonably well, but they did not appear to have sufficient attentional resources available to do both at once.

Brungart and Simpson's results suggest that tasks requiring a high attentional load can break down the binaural channel separation normally expected in dichotic listening. Significantly, when one speech masker was presented ipsilaterally, only a small additional effect on speech recognition at low S/Ms was found when a second masker composed of speech-shaped *noise* was presented contralaterally.

[a)]Portions of this work were presented at the 141st Meeting of the Acoustical Society of America [Kidd *et al.*, J. Acoust. Soc. Am. **109**, 2468(A) (2001)].
[b)]Electronic mail: gkidd@bu.edu

This latter result could mean that the contralateral masking effect only occurs for actual speech sounds and/or for the task of speech recognition.

The contralateral masking effect found in Brungart and Simpson's experiment was clearly due to limitations on central processing of the stimuli rather than to overlapping patterns of excitation in the auditory periphery. Auditory masking that results from central, rather than peripheral, processes has been termed ''informational masking'' (cf. Watson, 1987) and, with a few notable exceptions (e.g., Freyman *et al.*, 1999, 2001; Brungart, 2001; Brungart *et al.*, 2001; Arbogast *et al.*, 2002), has largely been studied using nonspeech stimuli in detection (e.g., Neff and Green, 1987; Neff, 1995; Kidd *et al.*, 1994; Oh and Lutfi, 1998; Wright and Saberi, 1999; Richards *et al.*, 2002) or discrimination/identification (e.g., Watson *et al.*, 1975, 1976; Howard *et al.*, 1984; Leek *et al.*, 1991; Neff and Jesteadt, 1996; Kidd *et al.*, 2002a) tasks. Although informational masking has been demonstrated for various nonspeech stimuli, it is not known whether the contralateral masking effect found by Brungart and Simpson is unique to speech recognition. There are no nonspeech results that we know of that appear directly comparable to those of Brungart and Simpson. Neff (1995) found only small threshold increases when an informational masker was presented in the ear contralateral to the pure-tone test signal in a detection task. However, much greater amounts of masking in a similar contralateral informational masking condition have been reported by Oh *et al.* (1999) in children. A plausible argument could be made that the contralateral masking effect found by Brungart and Simpson (2002) occurred because the listener is obliged to devote some amount of attentional resources to the processing of clearly audible speech even when it interferes with the task. The semantic content of speech may be particularly difficult to ignore, especially if there is some novel or inherently interesting information contained in the ''unattended'' message. According to that view, the nonspeech masker would not interfere with the recognition of the signal because it evoked no obligatory allocation of attention. Recently Conway *et al.* (2001) used a procedure in which a subject's own name was occasionally embedded in a stream of irrelevant speech presented to one ear (unattended ''contralateral'' ear) while attention was directed toward receiving information from the target in the opposite (attended ''ipsilateral'') ear (see also Moray, 1959 and Wood and Cowan, 1995). When the speech in the unattended ear was not novel or inherently interesting, no contralateral interference in the ipsilateral task was observed. However, about one-third of the subjects tested reported that they had heard their own name in the unattended ear and demonstrated performance deficits on the ipsilateral-ear task that coincided with the time period immediately after their name was presented. The implication is that the semantic content in the contralateral stimulus was important in producing the ipsilateral interference. Conway *et al.* (2001) also found that subjects who have relatively low working memory capacities were the most susceptible to the contralateral interference. Their findings could be construed as consistent with the hypothesis that the semantic content of speech is essential in breaking down the normally high degree of binaural channel separation or at least is an important contributor to that effect. However, Brungart and Simpson's results cannot be explained entirely by an argument based on the semantic content of speech. In one of their conditions, they used time-reversed speech as the contralateral masker. They found that time-reversed speech produced a contralateral masking effect comparable to that of normal speech, but only at negative signal-to-masker ratios. At higher signal-to-masker ratios, normal speech was a more effective masker. Thus it appears that, under certain circumstances, across-ear interference in speech recognition can occur from a contralateral masker that is sufficiently ''speechlike'' even in the absence of semantic content. However, it still does not indicate whether the contralateral informational masking effect is unique to the speech recognition task or to speech (or speechlike) maskers.

An alternative explanation for Brungart and Simpson's results is that the across-ear interference that was observed when the masker was added to the contralateral ear occurred because the signal and masker were qualitatively so similar that it was difficult for the listener to segregate the two sounds. Some studies of informational masking using nonspeech stimuli in detection or identification tasks have demonstrated that the similarity of the signal and masker can have a significant influence on the amount of informational masking that is produced (e.g., Kidd *et al.*, 1994, 2002a; Durlach *et al.*, 2003b). Thus, an alternative explanation is that the contralateral masking effect found by Brungart and Simpson reflects more general limitations on performance related to the ability to segregate sounds and is not unique to the speech recognition task or to maskers that are speech or speech-like.

The current experiments were designed to examine whether the contralateral masking effect found by Brungart and Simpson can be produced using nonspeech stimuli in a detection task. For that purpose, we used stimuli presented in several monotic and dichotic conditions that have been shown to produce varying degrees of informational masking while minimizing peripherally based energetic masking (Kidd *et al.*, 1994, 2002b).

II. METHODS

A. Subjects

A total of 18 subjects participated in this study. There were 12 subjects in experiment 1 and 8 subjects in experiment 2, with two subjects participating in both experiments. The subjects were college students ranging in age from 18 to 35 years. All had normal hearing as determined by audiometric evaluation. The experimental tasks were practiced for several hundred trials prior to collection of the data reported here.

B. Stimuli

The signal was a sequence of eight 60-ms (10/40/10 rise/steady-state/decay) bursts of a 1000-Hz tone. With the exception of one noise masker used in experiment 2 (described below) the maskers were also composed of eight 60-ms bursts played synchronously with the signal. In ex-

periment 1 there were three types of maskers: broadband noise (BBN), multiple-bursts same (MBS), and multiple-bursts different (MBD) informational maskers. The MBS and MBD maskers have been used previously in studies from our laboratory of informational masking (e.g., Kidd *et al.*, 1994, 2002b) and consist of sequences of random-frequency multitone complexes. The difference between MBS and MBD maskers is the way the randomization is applied. For MBS, a random set of frequencies is drawn for the first burst in a sequence, and the same frequencies are simply repeated in each successive burst. For MBD, each burst in a sequence is a different random draw of frequencies. The MBS and MBD maskers had eight equal-level components in each burst with an overall level of 70 dB SPL (approximately 61 dB SPL/component). The frequency range from which the MBS and MBD masker components were drawn was 200–5000 Hz with a protected region (no masker components) from 851 to 1175 Hz.

The overall level of the BBN masker was 70 dB SPL equaling about 33 dB spectrum level and was bandpass filtered from 200 to 5000 Hz. In experiment 2 the broadband noise was notch-filtered (two bands ranging from 200 to 600 Hz and 1400 to 1800 Hz) and played as synchronous bursts (SYN) or as a single constant burst during each observation interval (CON). In the CON case the masker onset was 200 ms prior to the onset of the signal and the masker offset was 200 ms following the offset of the signal. The SYN and CON noise maskers used in experiment 2 also had a spectrum level of 33 dB SPL and an overall level of about 62 dB SPL.

Typical examples of the two informational maskers and the BBN, SYN, and CON noise maskers are shown schematically in Fig. 1 as sound spectrograms. Note that the time scale is different for the CON masker than the other maskers. The left panels show representative draws of masker-only stimuli while the right panels show masker-plus-signal stimuli.

C. Procedures

The level of the signal was varied adaptively using a two-down one-up procedure that estimates the 70.7% correct point on the psychometric function (Levitt, 1971). Response feedback was provided after each trial. The sounds were played through matched TDH-50 earphones in a double-walled IAC booth. Response recording, interval timing, and response feedback were implemented using an LCD-display computer terminal.

All data shown are means from at least eight runs per subject. Each run consisted of at least 50 trials in which a minimum of nine reversals was obtained. The threshold estimates from each run were based on at least six reversals after discarding the first three or four reversals depending on whether the total number of reversals was odd or even. The order of conditions tested was mixed and randomized with the exception of one condition in experiment 2 as discussed below.

The signal was always in one ear designated the ''test ear.'' The subjects chose the test ear at the beginning of the study. The masker(s) could be played in the same ear as the signal (ipsilateral masker presentation) or in the opposite ear (contralateral masker presentation). When two maskers were presented they were always different types in the two ears.

III. EXPERIMENT 1

The first experiment was designed to test whether informational masking can be increased by the presentation of contralateral stimulation for the task of detecting the presence of a 1000-Hz signal. Table I contains a checklist of conditions tested. The rows of the table indicate the masking condition in the ipsilateral ear while the columns indicate the masking condition in the contralateral ear. There were three masking conditions in the test ear: signal in quiet (ipsilateral-NONE), MBS, and MBD. For each masking condition in the test ear there were three possible contralateral masking conditions: NONE, BBN, and whichever informational masker (MBS or MBD) was not present in the test ear. For the ipsilateral-NONE case all three maskers were tested contralaterally as was a condition where there was no masker in either ear.

Figure 2 shows the group mean thresholds and standard errors from all conditions from experiment 1. The abscissa is the masking condition in the ear contralateral to the signal. The three panels indicate the masking condition in the ear ipsilateral to the signal. The ordinate is signal threshold in dB SPL.

The left panel shows the results when there was no ipsilateral masker. With no masker in either ear, group mean threshold was about 1 dB SPL which serves as a reference for computing threshold shifts due to masking (also referred to as ''amount of masking''). Adding a masker to the contralateral ear shifted thresholds by about 3.5 to 6 dB. An analysis of variance indicated that this small effect of contralateral masker condition was statistically significant ($F[3,11]=20.3$, $p<0.001$). *Posthoc* analysis (Tukey's studentized range HSD test $p<0.05$) revealed that the BBN, MBS, and MBD maskers were all significantly different than the no-masker condition. The BBN and MBS maskers were also significantly different than MBD. This relatively small effect of a contralateral masker on quiet threshold in the test ear is consistent with the threshold shifts typically observed with contralateral maskers in clinical audiometry (e.g., Liden *et al.*, 1959), an effect which is referred to in the clinical literature as ''central masking'' (cf. Goldstein and Newman, 1985). There is also some study of the phenomenon in the scientific literature dating from Wegel and Lane (1924) and discussed in detail in Zwislocki *et al.* (1967). It is noteworthy that the BBN masker produced as much of a threshold shift as the MBS masker. It is possible that these small but statistically significant threshold shifts are due to the masker overcoming interaural attenuation and producing small amounts of masking in the ipsilateral ear. However, particularly for the informational maskers with the ''protected'' region around the signal frequency, it seems more likely that the effect is truly central in origin but not necessarily related to the informational masking strength of the masker.

In the ipsilateral-MBS conditions (middle panel), group mean threshold shifts (*re*: quiet threshold) were about 47 dB for the contralateral-NONE condition, 49 dB for the contralateral-MBD masker, and 53 dB for the contralateral-

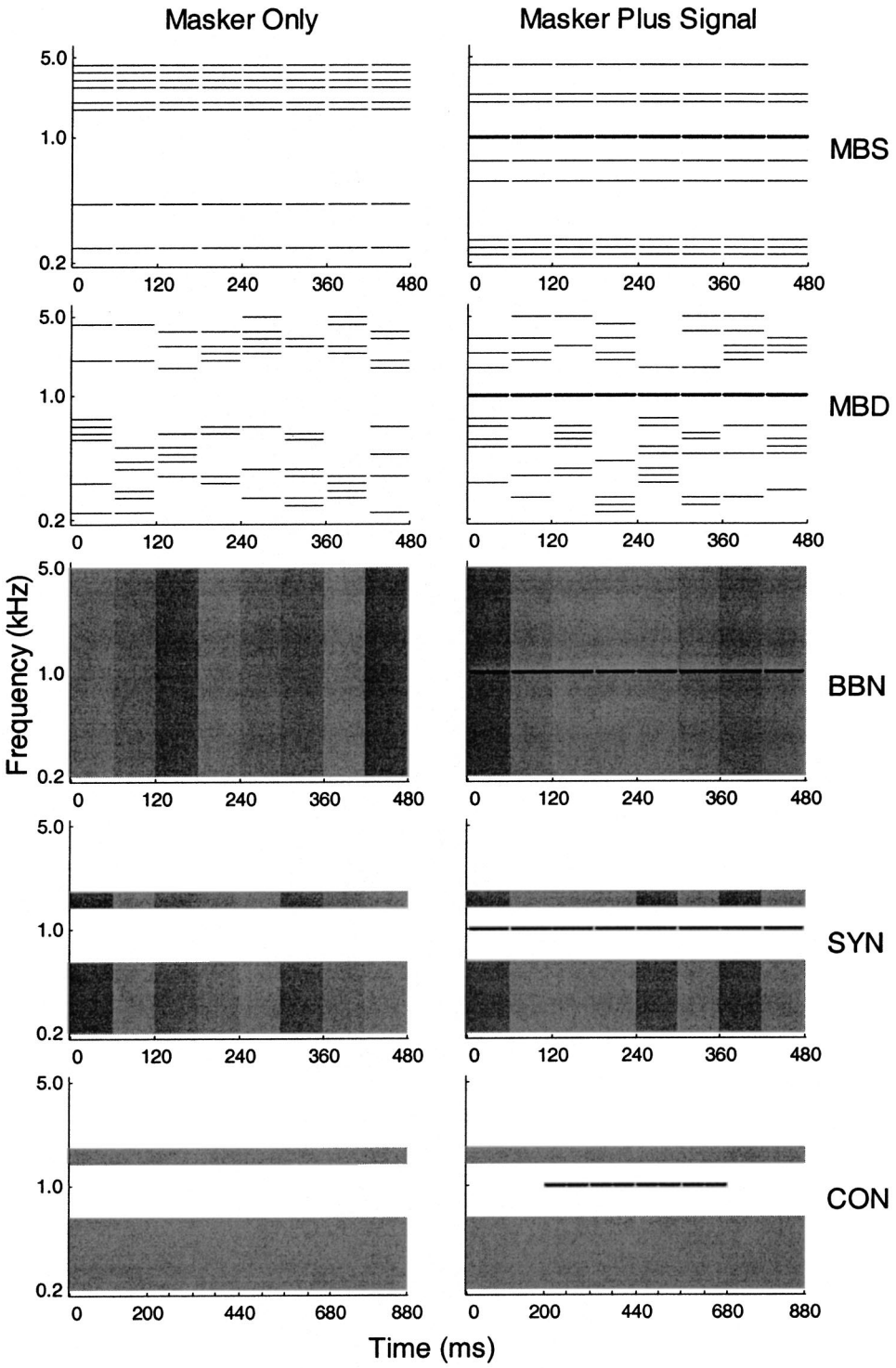


FIG. 1. A schematic of the maskers used in experiment 1 (upper three rows) and experiment 2 (lower two rows and upper two rows). The left panels show representative maskers alone and the right panels show representative maskers plus the signal. The upper row of panels illustrates the ''multiple-bursts same'' (MBS) masker, the second row of panels illustrates the ''multiple-bursts different'' (MBD) masker, the third row of panels illustrates the broadband noise (BBN) masker, the fourth row of panels illustrates the synchronously gated notched-noise (SYN) masker, and the final row of panels illustrates the constant notched-noise (CON) masker. Note that the abscissa (time) is the same for the top four rows of panels but is different in the final row of panels showing the CON masker.

BBN masker. The amount of masking produced by the ipsilateral-MBS masker is comparable to that reported in the past (cf. Kidd *et al.*, 1994) although some specifics of the experiments differ (e.g., number of bursts/burst duration). There were no statistically significant differences between the amount of masking produced by any of the three contralateral masking conditions. Thus, adding a contralateral masker in the ipsilateral-MBS condition did not significantly increase the amount of masking.

In the ipsilateral-MBD conditions (third panel), group mean threshold shifts were about 14.5 dB for the contralateral-NONE condition, 33.5 dB for the contralateral-MBS masker, and 21.3 dB for the contralateral-BBN masker. Thus the additional shift in ipsilateral-MBD masked threshold caused by adding a contralateral masker was about 19 dB for the MBS masker and about 7 dB for the BBN masker. The effect of contralateral masking condition was statistically significant ($F[2,11]=15.2$, $p<0.001$). *Posthoc* analysis (Tukey's studentized range HSD test $p<0.05$) indicated that the contralateral-MBS masker was significantly different than both the contralateral-BBN and contralateral-NONE conditions. The contralateral-BBN condition did not produce

TABLE I. Checklist of conditions tested in experiment 1. Note that the signal is always presented to the ipsilateral ear.

Masker in ipsilateral ear	Masker in contralateral ear			
	None	MBS	MBD	BBN
None	√	√	√	√
MBS	√		√	√
MBS	√	√		√

a statistically significant increase in masking *re*: the contralateral-NONE condition.

Figure 3 shows the individual thresholds in dB SPL from all conditions plotted as bar graphs. The abscissa in each panel is the contralateral masking condition from left to right: NONE, MBS, MBD, and BBN. The different panels (top to bottom) are for different ipsilateral masking conditions: NONE, MBS, and MBD. The individual subjects (shaded bars) are ordered from low to high based arbitrarily on the thresholds obtained in the ipsilateral-MBS contralateral-NONE condition. There are several aspects of the individual data that are of interest. First, as is typical of conditions producing large amounts of informational masking, thresholds for individual subjects were spread over a wide range for the conditions in which the MBS stimulus determined performance. This appeared to be the case for all conditions in which the MBS masker was ipsilateral to the signal, and also when the MBS masker was played contralateral to the signal in combination with the MBD masker presented ipsilaterally.

The dominant effect in this experiment was the presence of the MBS masker. When the MBS masker was played in the same ear as the signal, large amounts of masking occurred, on average, regardless of which masker (or even no masker) was present in the contralateral ear. The condition in which the masker in the ipsilateral ear was MBD while the masker in the contralateral ear was MBS provided the closest analog, we would argue, to the multiple-talker masking experiment of Brungart and Simpson (2002). In this case, the subjects apparently were unable to hold the inputs to the two ears separate and ignore the highly effective MBS masker in the contralateral ear. The amount of masking was not as great overall when the MBS masker was played contralaterally as when it was played ipsilaterally. These results are qualitatively similar to what might be expected if the MBS masker

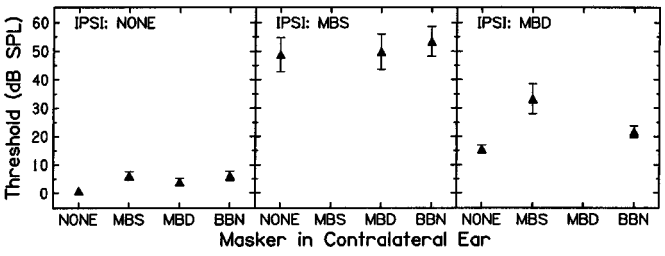


FIG. 2. Group mean thresholds in dB SPL and standard errors for all conditions of experiment 1. The abscissa is the masking condition in the ear contralateral to the signal. The left panel is for the signal alone in the test ear (ipsilateral-NONE), the middle panel is for the signal plus MBS masker in the test ear, and the right panel is for the signal plus MBD masker in the test ear.

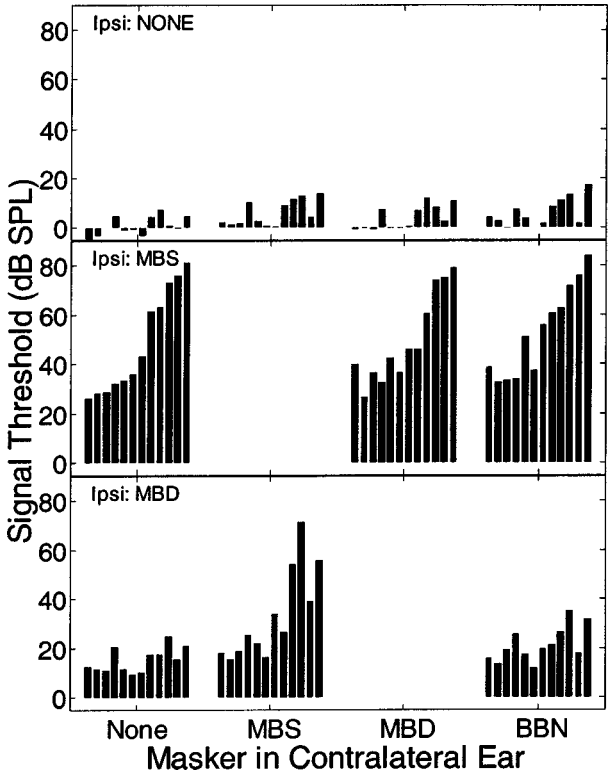


FIG. 3. Thresholds for individual subjects for all conditions of experiment 1. The abscissa is the masker presented in the ear contralateral to the signal. The upper panel is for the condition in which there was no masker in the test ear, the middle panel is when the MBS masker was presented in the test ear, and the lower panel is when the MBD masker was presented in the test ear. The subjects are ordered according to the amount of masking observed for the MBS-ipsilateral only condition (left graph middle panel). That same subject order is preserved in all panels shown.

were simply attenuated and presented ipsilaterally to the test ear. However, as was the case in the Brungart and Simpson study, there must be another sound present in the test ear besides the signal for the contralateral masker to be effective. Thus, the contralateral masking effect reported by Brungart and Simpson for multitalker speech recognition does not appear to be specific to the use of speech as a stimulus. Nor is the breakdown of binaural channel separation implied by this result specific to the speech recognition task; it also can be quite a strong effect for nonspeech detection tasks. Further, the results from this experiment support the finding of Brungart and Simpson that the composition of the contralateral masker is critical in producing the effect. When the contralateral masker was broadband noise a relatively small and statistically insignificant increase in threshold was observed.

IV. EXPERIMENT 2

The second experiment was designed to examine the role played by the ipsilateral masker in producing the contralateral masking effect. It is clear from experiment 1 that the contralateral masking effect occurs for an ipsilateral MBD masker combined with a contralateral MBS masker. And, although there was a statistically significant threshold increase for the ipsilateral-NONE condition with any contralateral masker, the effects were much smaller. Because this effect was not sensitive to contralateral masker type in the same way as for the ipsilateral-MBD condition, this probably

means that the two effects are not closely related. That raises the question of whether the contralateral masking effect would occur for *any* ipsilateral masker or whether there is something special about the MBD masker that causes the listener to be susceptible to the contralateral stimulus. No contralateral masking effect was found when the ipsilateral masker was MBS. However, the MBS masker was not paired with an effective contralateral informational masker and, given the amount of masking produced in the ipsilateral-MBS only condition, may have overwhelmed any contralateral masking influence present. In order to begin to answer the question of the role of the ipsilateral masker, we tested three different types of ipsilateral maskers paired with the contralateral MBS masker. The three ipsilateral maskers were MBD, notch-filtered Gaussian noise gated synchronously with the signal throughout the burst sequence (SYN), and notch-filtered Gaussian noise that was gated on prior to the signal and gated off following the signal but which was continuous during the observation interval (CON). The reason for testing noise as an ipsilateral masker is that it tends to produce largely energetic masking rather than informational masking which may be an important ipsilateral factor in producing the contralateral masking effect. The reason for testing gated versus continuous noise is that the synchronous gating may also be important in producing the effect because synchronous onsets and offsets are known to be strong grouping cues while asynchronous onsets and offsets promote perceptual segregation (e.g., Darwin and Carlyon, 1995). The noise maskers were notched filtered to provide an analog of the ''protected region'' used in generating the MBD masker. The MBS and MBD stimuli were the same as in experiment 1. These maskers are also represented graphically in Fig. 1.

Eight subjects participated in this experiment. Two of the subjects had participated in experiment 1 and the two Gaussian noise conditions—SYN and CON—were completed after experiment 1. For the other six subjects, the order of testing the conditions was such that the ipsilateral-only conditions were tested first, then the contralateral-MBS conditions with the three different ipsilateral maskers were tested in mixed and randomized order.

It should be pointed out that the comparison of the amount of masking obtained ipsilaterally for the MBD masker and the two noise maskers should be viewed with some caution. The intent was to choose masker values that would produce roughly the same ipsilateral masked thresholds so that a reasonably fair comparison could be made of the effect of adding the contralateral MBS masker. The noise maskers had the same spectrum level as the BBN masker in experiment 1 (which was equal in SPL to the MBD and MBS maskers) although, due to the differences in filtering, the SPL for CON and SYN was about 8 dB less. However, we do not know how the susceptibility to the contralateral masker depends on the amount of masking in the ipsilateral ear for these maskers so it is possible that there is an interaction with amount of masking or masker type that would not be apparent here.

Table II contains a checklist of conditions tested in experiment 2. The amount of masking (group means *re*: quiet threshold) produced by the ipsilateral maskers, with no contralateral masker, was 26 dB for the CON masker, 25.8 dB for the SYN masker, and 19.7 dB for the MBD masker. What is of interest here is not how much masking was produced by the ipsilateral maskers but rather how much more masking occurred when the contralateral MBS masker was added. This *additional* masking due to the presence of the contralateral masker is the *contralateral masking effect* of importance for these stimuli and conditions. Presentation of the contralateral MBS masker when there was no ipsilateral masker produced about 6 dB of masking which was very similar to that which was found in experiment 1. Presentation of the MBS masker contralaterally caused additional shifts in threshold of 7.3, 9.9, and 18.7 dB for the three ipsilateral maskers, respectively. The additional shifts in threshold produced by the contralateral MBS masker, relative to the amounts of masking produced by the ipsilateral maskers alone, are plotted in Fig. 4 for those ipsilateral maskers and for the ipsilateral-NONE (contralateral MBS masker, no masker in signal ear) condition. An analysis of variance of the results shown in Fig. 4 indicated that masker type was a significant effect ($F[3,21]=7.60$, $p<0.005$). *Posthoc* analyses (Tukey's studentized range HSD test, $p<0.05$) indicated that there was no significant difference between the ipsilateral-NONE, ipsilateral-CON, and ipsilateral-SYN con-

TABLE II. Checklist of conditions tested in experiment 2 (asterisks indicate conditions that are in common with experiment 1). Note that the signal is always presented to the ipsilateral ear.

Masker in ipsilateral ear	Masker in contralateral ear	
	None	MBS
None	√*	√*
MBD	√*	√*
SYN	√	√
CON	√	√

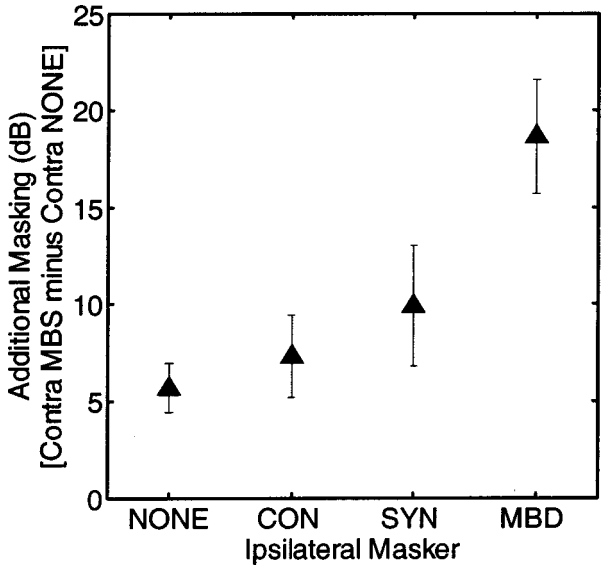


FIG. 4. Group mean results from experiment 2. The ordinate is the amount of additional masking produced by presenting the MBS masker to the ear contralateral to the signal and the abscissa is the masking condition in the ipsilateral ear. The error bars are standard errors of the means.

ditions while the ipsilateral-MBD condition was different than the other three conditions. What this means is that the contralateral MBS masker produced no more additional masking for the SYN or CON maskers than it did for no ipsilateral masker. However, when the ipsilateral masker was MBD, a statistically greater amount of additional masking was found. Thus, it is not the presence of a masker *per se* in the ipsilateral ear that makes the listener especially susceptible to contralateral informational masking, but it appears that it is the characteristics of the ipsilateral masker, or the processing of the ipsilateral masker required to solve the task, that matters. Taking into account the results from experiments 1 and 2, it seems that the properties of *both* the ipsilateral and contralateral maskers are crucial for producing the contralateral masking effect found here, and likely that found by Brungart and Simpson for multiple talkers. Recently, Brungart and Simpson (2002) reported a similarly reduced effect of contralateral informational masking when the ipsilateral masker was noise rather than speech and arrived at a similar conclusion.

V. CONDITIONS COMMON TO BOTH EXPERIMENTS 1 AND 2

All 18 subjects performed ipsilateral MBS and MBD masked conditions as well as signal in quiet. Further, all of the subjects were tested with the MBS masker played contralateral to the MBD masker and to the signal in quiet. For the entire group, the average threshold shift caused by the ipsilateral-MBD masker alone was 16.9 dB while the shift caused by the ipsilateral-MBS masker alone was 49.6 dB. The correlation across subjects between the amount of masking for those two conditions was not statistically significant ($r=0.34$, $p>0.05$). For the condition which combined the ipsilateral-MBD masker with the contralateral-MBS masker the average threshold shift was 34.3 dB. Further, there was a positive correlation between the amount of threshold shift caused by the MBS masker played ipsilateral-only and when played contralaterally with the MBD masker ipsilateral. This point is illustrated by a plot of the individual results in Fig. 5 (also apparent to a lesser degree in Fig. 3). Figure 5 shows the individual threshold shifts (*re*: quiet threshold) for the condition ipsilateral-MBD contralateral-MBS as a function of the thresholds for ipsilateral-MBS only (abscissa). The line is a least-squares fit (slope=0.63, $r=0.73$, $p<0.001$). The significant correlation between the thresholds in these two conditions suggests that a subject who tends to be susceptible to informational masking ipsilaterally will also be susceptible when the masker is played contralaterally. Thus, whether an informational masker affects a subject is largely unrelated to whether the masker is presented in the same ear as the signal or in the contralateral ear, so long as there is a sufficiently complex stimulus and/or task to be performed in the test ear. It should be noted, though, that contralateral presentation of MBS with ipsilateral-MBD was not quite as effective as ipsilateral presentation of the MBS masker.

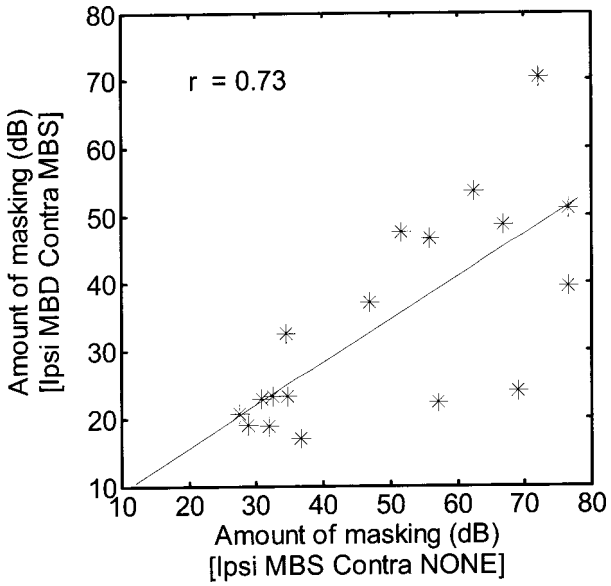


FIG. 5. Individual results from all 18 subjects for two conditions tested in both experiments 1 and 2. The abscissa indicates the amount of masking produced by the MBS ipsilateral-only condition while the ordinate indicates the amount of masking in the condition where the MBD masker is played in the ipsilateral ear and the MBS masker is played in the contralateral ear. The solid line is a linear least-squares fit to the data.

VI. DISCUSSION

The discussion will center on three main points. First, the phenomenon reported by Brungart and Simpson in which the binaural separation of channels breaks down in certain dichotic listening conditions does not appear to be unique to speech stimuli or the speech recognition task. Second, the degree of binaural channel separation is variable and depends on the stimulus and/or task in *both* ears. And third, there seems to be something important about the ipsilateral-MBD stimulus and task that leads to the breakdown of binaural channel separation. The latter must be understood if the phenomenon is to be fully explained. Included in this understanding is the relation to the ipsilateral speech signal and masker used by Brungart and Simpson that apparently led to the same type of breakdown in binaural channel separation. These three points will be discussed in order.

With respect to the original hypothesis that the contralateral masking effect found by Brungart and Simpson is unique to speech or to the speech recognition task, the present results seem quite clear. It is possible using nonspeech sounds and the task of detection to construct listening conditions where, for many subjects, the high degree of binaural channel separation normally observed breaks down. Thus, the contralateral masking effect does not depend on the stimulus having semantic content. This is true for the target, the ipsilateral masker, and the contralateral masker. However, our results do not preclude the possibility that the amount of across-ear masking increases when the stimulus or the task becomes more complex, in which case multitalker speech recognition conditions may represent one of the more challenging situations for maintaining binaural channel separation. Brungart and Simpson (2002) found that the contralateral masking effect was present, but smaller, when the ipsilateral speech masker was time reversed relative to the

case where it was presented normally. Presumably, the spectrotemporal complexity of time-reversed speech would be approximately the same as normal speech but the linguistic content would be eliminated. The normal speech was a more effective masker particularly at high signal-to-noise ratios where, as Brungart and Simpson speculated, the lack of linguistic content of the time-reversed speech would make the task of segregating the target easier. The finding that time-reversed speech contributes to the susceptibility to contralateral interference, but does so less than normal speech, suggests that the processing of the semantic content of speech can be a factor in the extent to which binaural channels can be held separate.

With respect to the second point, the current results support those of Brungart and Simpson (2002) which indicate that the extent to which listeners can hold the inputs from the two ears separate is not fixed but depends on events occurring in both ears. Also consistent with the findings of Brungart and Simpson are the current results indicating that greater interference occurs for an informational masker presented contralaterally than for an energetic masker presented contralaterally. This finding may be viewed as consistent with the distinction usually drawn between the two types of masking. Energetic masking is thought of as overlapping patterns of excitation on the basilar membrane and in the auditory nerve. Thus, energetic masking occurs prior to the first physiological site receiving inputs from the two ears. Informational masking, in contrast, is usually thought of as occurring despite a sufficiently robust neural representation of the signal, at least at the level of the auditory nerve, for the listener to solve the task. Thus, it is higher level processing that breaks down and leads to masking. Given that explanation for the two types of masking, it is not surprising that the informational masker has a much greater effect than the energetic masker when presented to the contralateral ear. A related point concerning the nonperipheral nature of the contralateral effect found here is the significant correlation between the amount of masking found in the ipsilateral-MBS contralateral-NONE condition and the ipsilateral-MBD contralateral-MBS condition. Although the possibility that some of the large intersubject differences found in informational masking are peripherally based has perhaps not received the consideration it deserves (cf. Durlach *et al.*, 2003a), the results shown in Fig. 5 would seem inconsistent with peripheral factors—as we have discussed them here—playing much of a role in these conditions.

It is interesting to note that our results differ from those of Brungart and Simpson in that the intersubject differences in susceptibility to contralateral masking were much greater here. This likely is due to the differences in experimental tasks. Listeners with normal hearing and language obviously engage in speech recognition regularly as part of their daily lives and might not be expected to differ in performance of such tasks to the extent that they would vary in detection of a sequence of pure tones. In fact, large individual differences in informational masking for pure tone detection tasks have been reported and commented on previously (e.g., Neff *et al.*, 1993; Neff and Dethlefs, 1995; Durlach *et al.*, 2003b) and may be related to a variety of factors such as musical training and past listening experience (e.g., Oxenham *et al.*, 2003).

With respect to the third point made at the beginning of this discussion, it is of interest to determine which characteristics of the stimulus in the ipsilateral ear, or the associated task, contribute to the breakdown of the channel separation between ears. Clearly, it is the ipsilateral-MBD stimulus and/or task that modulates the susceptibility to the masker from the contralateral channel. If it were simply a matter of the signal fusing perceptually with the MBS masker because of similar spectrotemporal properties (i.e., constant-frequency coherently amplitude-modulated tones), then we would have expected that the control condition of ipsilateral-NONE contralateral-MBS would produce much greater masking than that which was observed. Or, because of the greater envelope coherence, we would have expected the ipsilateral-SYN condition to be more susceptible to the contralateral-MBS masker than the ipsilateral-CON condition. Clearly, though, when the ipsilateral masker is noise, the effect of the contralateral MBS masker is greatly reduced. One possible explanation concerns the level at which auditory stream formation occurs relative to the combination—or selective exclusion—of binaural information. It seems plausible that the observer must devote greater attention to the test ear when the signal is embedded in the MBD masker than when it is in quiet or in noise. Perceptually, in MBD, the signal tone forms a coherent stream as the burst sequence progresses eventually segregating from the randomly varying masker over time. Perhaps it is the processing that is required to form the signal stream in MBD that imposes greater demands on the listener than merely detecting a tone in quiet or in a qualitatively dissimilar noise. Recently, Carlyon *et al.* (2001) have demonstrated that the buildup of streaming in one ear could be greatly reduced by having the listener direct attention to a task in the contralateral ear. Further, for patients with brain damage causing hemispheric neglect, stream segregation was greatly reduced in the affected ear but not in the contralateral, unaffected ear. Their results indicate that stream formation occurs as a top-down process that depends on allocation of attention. In a recent study from our laboratory (Kidd *et al.*, 2002b), it was found that temporally separating the bursts in the MBD stimulus increased thresholds considerably. Although there were predictably large differences between subjects, Kidd *et al.* found that increasing the interburst interval (IBI) from 0 to 400 ms increased group-mean masked thresholds in an eight-burst MBD masker by about 25 dB. The explanation for this large effect was that increasing the IBI disrupted the coherence of the signal tone so that it no longer perceptually segregated from the masker bursts. ''Hearing out'' the signal tone in the MBD stimulus subjectively seems (to us) effortless and the perception of an auditory stream separate from the masker is quite salient. However, it may well be that the perceptual formation of an auditory stream requires substantial allocation of attentional resources which appears to be an important factor in holding the binaural channels separate.

In Brungart and Simpson's (2002) work, one piece of evidence that the masking that occurred was informational masking is that a high proportion of the errors made in the

speech recognition task actually were confusions with the masker. That is, the colors and numbers that were erroneously reported by the subjects were those from the masker sentences and were not random as might be expected if the subjects were guessing. In the case where there were two speech maskers—one ipsilateral and one contralateral—Brungart and Simpson found the great majority of confusions came from the ipsilateral masker sentences and not the contralateral masker sentences. They interpreted that result as indicating that the failure in the segregation task was in the ipsilateral ear. Our results do not really permit a similar determination of a presumed failure of segregation in one ear or the other. We have interpreted the lesser amount of masking found for the MBD masker compared to a single-burst masker (one random draw of frequencies played as a single burst in the observation interval; e.g., Neff and Green, 1987) as being a result of the segregation of the constant-frequency signal from the randomly varying masker (Kidd *et al.*, 1994, 2002b). So, the argument could be made that, analogous to the conclusion reached by Brungart and Simpson for their stimuli, the contralateral masking effect found for ipsilateral-MBD/contralateral-MBS was due primarily to a failure to segregate the signal from the MBD masker in the ipsilateral ear. However, we should point out that, as shown in Fig. 5, the significant correlation between the ipsilateral-MBS thresholds and the ipsilateral-MBD/contralateral-MBS thresholds strongly suggest that it was the MBS masker that was responsible for the informational masking that was observed regardless of the ear to which it was presented. Thus, we cannot rule out the possibility that the ipsilateral-MBD masker simply allowed the contralateral-MBS masker to be ''mixed'' in the higher level representation of the stimulus.

A final point concerns speculation about the extent to which the breakdown in channel separation found here and in Brungart and Simpson's work happens when sounds are played in real rooms rather than through headphones. It is probably accurate to conclude that binaural channel separation is neither a complete nor a natural condition of hearing. We do not have ''earlids'' that allow us to block acoustic input to one channel as eyelids block optical input to the visual system. In addition to the work here and that of Brungart and Simpson (2002), there are other examples of instances where binaural channel separation fails. The study by Conway *et al.* (2001) discussed above regarding performance deficits occurring for an attended ear due to presentation of inherently interesting information in the contralateral unattended ear is one such example. Another example was reported by Heller and Trahiotis (1995). In their study, four noise bursts were presented in sequence to one ear of the listener. Three of the bursts were identical while one of the two middle bursts was a different sample with the task being to choose which of the two middle samples was different. As has been shown previously, discriminability of different noise bursts can be quite good for certain combinations of noise bandwidth and duration (e.g., Hanna, 1984). Compared to this monaural condition, however, Heller and Trahiotis found that presentation of uncorrelated synchronous noise bursts to the contralateral ear could greatly decrease discrimination performance. The implication is that, despite the fact that it would be of great benefit for the listener to ignore the ear contralateral to the test stimuli, the listeners were unable to do so.

Normally, binaural information is combined in useful ways to locate images in the environment providing spatial ''markers'' for the focus of attention or, perhaps, to ''null out'' locations to which we do not wish to attend (e.g., Durlach *et al.*, 2003a; also, related work by Arbogast and Kidd, 2000; Kidd *et al.*, 1998). Although we cannot be certain, the current results imply that the direction-dependent attenuation of unwanted acoustic information is the complex interaction of the difficulty in solving the task at the attended location and the distraction caused by unwanted information from other locations.

To summarize, the main findings of this study were the following.

(1) On average, when the signal was played in quiet, small shifts in detection threshold (3.5–6 dB) occurred when any masker was presented to the contralateral ear.
(2) When the ipsilateral masker was MBS, large shifts in group mean thresholds for detecting the signal (*re*: quiet threshold) occurred that were not significantly elevated when the MBD or BBN maskers were added to the contralateral ear. Large differences in the amount of masking were observed across subjects.
(3) When the ipsilateral masker was MBD, masked thresholds were relatively low when there was no contralateral masker and were not increased significantly by adding the contralateral-BBN masker. However, when the ipsilateral masker was MBD, adding the MBS masker contralaterally significantly increased the amount of masking.
(4) There was a significant positive correlation between the amount of masking of the signal in the ipsilateral-MBS contralateral-NONE condition and the amount of masking of the signal in the ipsilateral-MBD contralateral-MBS condition. This indicated that, for subjects who tended to be susceptible to informational masking, it did not greatly matter whether the informational masker was presented in the same ear as the signal or in the opposite ear as long as there was a difficult sound segregation task in the ipsilateral ear.
(5) The effect of the contralateral masker depended on the ipsilateral masker. Much less of an effect of the MBS masker played contralaterally was observed when the ipsilateral masker was Gaussian noise than when the ipsilateral masker was MBD.

ACKNOWLEDGMENTS

This work was supported by NIH/NIDCD Grant Nos. DC04545 and DC00100 and by the Boston University Hearing Research Center. The authors are grateful to Leslie R. Bernstein, Andrew J. Oxenham, and an anonymous reviewer for comments on an earlier draft of this manuscript.

Arbogast, T. L., and Kidd, Jr., G. (**2000**). ''Evidence for spatial tuning in informational masking using the probe-signal method,'' J. Acoust. Soc. Am. **108**, 1803–1810.

Arbogast, T. L., Mason, C. R., and Kidd, Jr., G. (**2002**). ''The effect of spatial separation on informational and energetic masking of speech,'' J. Acoust. Soc. Am. **112**, 2086–2098.
Broadbent, D. E. (**1958**). *Perception and Communication* (Pergammon, London).
Bronkhnorst, A. W. (**2000**). ''The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,'' Acustica **86**, 117–128.
Brungart, D. S. (**2001**). ''Informational and energetic masking effects in the perception of two simultaneous talkers,'' J. Acoust. Soc. Am. **109**, 1101–1109.
Brungart, D. S., and Simpson, B. D. (**2002**). ''Within-ear and across-ear interference in a cocktail party listening task,'' J. Acoust. Soc. Am. **112**, 2985–2995.
Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). ''Informational and energetic masking effects in the perception of multiple simultaneous talkers,'' J. Acoust. Soc. Am. **110**, 2527–2538.
Carhart, R., Tillman, T. W., and Greetis, E. S. (**1959**). ''Perceptual masking in multiple sound background,'' J. Acoust. Soc. Am. **45**, 694–702.
Carlyon, R. P., Cusack, R., Foxton, J. M., and Robertson, I. H. (**2001**). ''Effects of attention and unilateral neglect on auditory stream segregation,'' J. Exp. Psychol. Hum. Percept. Perform. **27**, 115–127.
Cherry, E. C. (**1953**). ''Some experiments on the recognition of speech, with one and two ears,'' J. Acoust. Soc. Am. **25**, 975–979.
Conway, A. R. A., Cowan, N., and Bunting, M. F. (**2001**). ''The cocktail party phenomenon revisited: The importance of working memory capacity,'' Psychonom. Bull. Rev. **8**, 331–335.
Darwin, C. J., and Carlyon, R. P. (**1995**). ''Auditory Grouping,'' in *Hearing: from Handbook of Auditory Perception*, 2nd ed., edited by B. C. J. Moore (Academic, London), pp. 387–424.
Durlach, N. I., Mason, C. R., Kidd, Jr., G., Arbogast, T. L., Shinn-Cunningham, B. C., and Colburn, H. S. (**2003a**). ''Note on informational masking'' J. Acoust. Soc. Am. (in press).
Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Coburn, H. S., and Kidd, Jr., G. (**2003b**). ''Information masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity'' (submitted for publication).
Feston, J. M., and Plomp, R. (**1990**). ''Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing,'' J. Acoust. Soc. Am. **88**, 1725–1736.
Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). ''The role of perceived spatial separation in the unmasking of speech,'' J. Acoust. Soc. Am. **106**, 3578–3588.
Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). ''Spatial release from informational masking in speech recognition,'' J. Acoust. Soc. Am. **109**, 2112–2122.
Goldstein, B. A., and Newman, C. W. (**1985**). ''Clinical masking: A decision-making process,'' in *Handbook of Clinical Audiology*, 3rd ed., edited by J. Katz (Williams and Wilkins, Baltimore), pp. 170–201.
Hanna, T. E. (**1984**). ''Discrimination of reproducible noise as a function of bandwidth and duration,'' Percept. Psychophys. **36**, 409–416.
Heller, L. M., and Trahiotis, C. (**1995**). ''The discrimination of samples of noise in monotic, diotic, and dichotic conditions,'' J. Acoust. Soc. Am. **97**, 3775–3781.
Howard, J. H., O'Toole, A. J., Parasuraman, R., and Bennett, K. B. (**1984**). ''Pattern-directed attention in uncertain-frequency detection,'' Percept. Psychophys. **35**, 256–264.
Kidd, Jr., G., Mason, C. R., and Arbogast, T. L. (**2002a**). ''Similarity, uncertainty and masking in the identification of nonspeech auditory patterns,'' J. Acoust. Soc. Am. **111**, 1367–1376.
Kidd, Jr., G., Mason, C. R., and Richards, V. M. (**2002b**). ''Measuring informational masking using the multiple-bursts procedure: Effects of number of bursts and interburst interval,'' Program of the 24th Midwinter Research Meeting of the Association for Research in Otolaryngology.
Kidd, Jr., G., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**). ''Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns,'' J. Acoust. Soc. Am. **104**, 422–431.
Kidd, Jr., G., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). ''Reducing informational masking by sound segregation,'' J. Acoust. Soc. Am. **95**, 3475–3480.
Leek, M. R., Brown, M. E., and Dorman, M. F. (**1991**). ''Informational masking and auditory attention,'' Percept. Psychophys. **50**, 205–214.
Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Liden, G., Nilssen, G., and Anderson, H. (**1959**). ''Masking in clinical audiometry,'' Acta Otolaryngol. (Stockholm) **50**, 125–136.
Moray, N. (**1959**). ''Attenuation in dichotic listening: Affective cues and the influence of instructions,'' Q. J. Exp. Psychol.: General **126**, 211–227.
Neff, D. L. (**1995**). ''Signal properties that reduce masking by simultaneous random-frequency maskers,'' J. Acoust. Soc. Am. **98**, 1909–1920.
Neff, D. L., and Dethlefs, T. M. (**1995**). ''Individual differences in simultaneous masking with random-frequency, multicomponent maskers,'' J. Acoust. Soc. Am. **98**, 125–134.
Neff, D. L., and Green, D. M. (**1987**). ''Masking produced by spectral uncertainty with multicomponent maskers,'' Percept. Psychophys. **41**, 409–415.
Neff, D. L., and Jesteadt, W. (**1996**). ''Intensity discrimination in the presence of random-frequency, multicomponent maskers and broadband noise,'' J. Acoust. Soc. Am. **100**, 2289–2298.
Neff, D. L., Dethlefs, T. M., and Jesteadt, W. (**1993**). ''Informational masking for multicomponent maskers with spectral gaps,'' J. Acoust. Soc. Am. **94**, 3112–3126.
Oh, E., and Lutfi, R. A. (**1998**). ''Nonmonotonicity of informational masking,'' J. Acoust. Soc. Am. **104**, 3489–3499.
Oh, E., Wightman, F. L., and Lutfi, R. A. (**1999**). ''The effect of spatial separation on informational masking in preschool children,'' 22nd Midwinter Research Meeting, Assoc. Res. Otolaryn.
Oxenham, A. J., Fligor, B., Mason, C. R., and Kidd, Jr., G. (**2003**). ''Informational masking and musical training'' (submitted for publication).
Richards, V. M., Tang, Z., and Kidd, Jr., G. (**2002**). ''Informational masking with small set sizes,'' J. Acoust. Soc. Am. **111**, 1359–1366.
Watson, C. S. (**1987**). ''Uncertainty, informational masking and the capacity of immediate auditory memory,'' in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 267–277.
Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). ''Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of uncertainty,'' J. Acoust. Soc. Am. **60**, 1176–1186.
Watson, C. S., Wroton, H. W., Kelly, W. J., and Benbasset, C. A. (**1975**). ''Factors in the discrimination of tonal patterns. I. Component frequency, temporal position and silent intervals,'' J. Acoust. Soc. Am. **57**, 1175–1185.
Wegel, R. L., and Lane, C. E. (**1924**). ''The auditory masking of one pure tone by another and its probable relationship to the dynamics of the inner ear,'' Physiol. Rev. **23**, 266–285.
Wood, N. L., and Cowan, N. (**1995**). ''The cocktail party phenomenon revisited: Attention and memory in the classic selective listening procedure of Cherry (1951),'' J. Exp. Psychol. Learn Mem. Cogn. **21**, 255–260.
Wright, B. A., and Saberi, K. (**1999**). ''Strategies used to detect auditory signals in small sets of random maskers,'' J. Acoust. Soc. Am. **105**, 1765–1775.
Yost, W. A. (**1997**). ''The cocktail party problem: Forty years later,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. A. Gilkey and T. R. Anderson (Erlbaum, NJ), pp. 329–348.
Zwislocki, J. J., Damianopoulos, E. N., Buining, E., and Glanz, J. (**1967**). ''Central masking: Some steady-state and transient effects,'' Percept. Psychophys. **2**, 59–64.

Spectral shape discrimination by hearing-impaired and normal-hearing listeners

Jennifer J. Lentz[a] and Marjorie R. Leek
Army Audiology and Speech Center, Walter Reed Army Medical Center, Washington, DC 20307

(Received 12 December 2001; revised 12 December 2002; accepted 16 December 2002)

The ability to discriminate between sounds with different spectral shapes was evaluated for normal-hearing and hearing-impaired listeners. Listeners discriminated between a standard stimulus and a signal stimulus in which half of the standard components were decreased in level and half were increased in level. In one condition, the standard stimulus was the sum of six equal-amplitude tones (equal-SPL), and in another the standard stimulus was the sum of six tones at equal sensation levels *re*: audiometric thresholds for individual subjects (equal-SL). Spectral weights were estimated in conditions where the amplitudes of the individual tones were perturbed slightly on every presentation. Sensitivity was similar in all conditions for normal-hearing and hearing-impaired listeners. The presence of perturbation and equal-SL components increased thresholds for both groups, but only small differences in weighting strategy were measured between the groups depending on whether the equal-SPL or equal-SL condition was tested. The average data suggest that normal-hearing listeners may rely more on the central components of the spectrum whereas hearing-impaired listeners may have been more likely to use the edges. However, individual weighting functions were quite variable, especially for the HI listeners, perhaps reflecting difficulty in processing changes in spectral shape due to hearing loss. Differences in weighting strategy without changes in sensitivity suggest that factors other than spectral weights, such as internal noise or difficulty encoding a reference stimulus, also may dominate performance. [DOI: 10.1121/1.1553461]

PACS numbers: 43.66.Fe, 43.66.Sr, 43.66.Jh [NFV]

I. INTRODUCTION

Over the last 20 years, a number of experiments have evaluated how the auditory system distinguishes between sounds with different spectral shapes, or profiles (Green, 1988). In these ''profile-analysis'' experiments the stimuli often are the sum of tones, and a signal stimulus with a different spectral shape is generated by altering the level of one or more tones of a standard stimulus. Stimuli with tones sparsely spaced in frequency are commonly used so that excitation produced by one tone does not overlap greatly with excitation produced by another (Green and Mason, 1985; Bernstein and Green, 1987). In profile-analysis experiments, the overall levels of the stimuli are varied on every presentation to reduce the possibility that listeners discriminate between these sounds based on a change in level of a single stimulus component. A large range of level variation (rove) typically is used to ensure that thresholds for changes in spectral shape are below the prediction made if listeners were making decisions based on the overall level of the stimuli or the level of a single component. In the presence of large level variation, listeners must compare levels across frequency within each stimulus to discriminate between two sounds with different power spectra (Green, 1988).

Correlation methods can be used to evaluate underlying decision strategies employed by the auditory system in discriminating between sounds with differing spectral shapes. These methods allow an experimenter to estimate the relative importance (or relative weight) that a particular stimulus component has on the discrimination of these sounds (e.g., Berg, 1989, 1990; Berg and Green, 1990; Richards and Zhu, 1994; Lutfi, 1995). When applying these methods, an experimenter alters the levels of individual stimulus components randomly and independently every time a stimulus is presented to a listener. In a two-interval forced-choice task, the listener's response (signal present in interval 1 or 2) can be correlated with the level difference between the two intervals for each stimulus component. Results from this analysis will estimate each component's contribution to the listener's decision and can be compared with predictions of an ideal observer. In many instances, normal-hearing listeners adopt an optimal decision strategy when discriminating between profile-analysis stimuli (Berg and Green, 1990; Dai *et al.*, 1996).

It is well established that normal-hearing listeners can easily compare the spectral shapes of sounds under conditions where the spectral shape change is not substantially altered internally by frequency selectivity. This ability has been tested to a much lesser extent in hearing-impaired listeners. Kidd and Mason (1992) pointed out that overall level variation potentially leads to experimental confounds when measuring across-frequency integration in hearing-impaired listeners. Hearing loss configurations may interact with roving levels: at low stimulus levels a sloping hearing loss acts as a low-pass filter, resulting in a stimulus where high-frequency components are inaudible. At high stimulus intensities, recruitment in impaired ears may cause sounds to be

[a]Current affiliation: Department of Speech and Hearing Sciences, Indiana University, 200 S. Jordan Ave., Bloomington, IN 47405. Electronic mail: jjlentz@indiana.edu

uncomfortably loud. Doherty and Lutfi (1996, 1999) circumvented many of these experimental confounds by studying across-frequency integration of level information in intensity discrimination tasks. In 1996, Doherty and Lutfi asked normal-hearing and hearing-impaired listeners to discriminate broadband intensity changes of stimuli that were the sum of tones. Their data indicated that hearing-impaired listeners relied primarily on components near the edge of their hearing loss. Their 1999 follow-up study also supported this conclusion.

The data obtained by Doherty and Lutfi (1996, 1999) suggest that decision strategies can and do change with cochlear damage, but these conclusions are based on intensity discrimination tasks only. Whether these results will generalize to a broader range of tasks has yet to be determined. Demands on the auditory system may be different for intensity discrimination and spectral shape discrimination tasks. For example, intensity discrimination is affected by the duration of the interstimulus interval (ISI), whereas profile-analysis is not (Green *et al.*, 1983). Lentz and Richards (1997) were unable to jointly account for broadband intensity discrimination and spectral shape discrimination data using an optimal, linear model (Durlach *et al.*, 1986). Lentz and Richards posited that the presence of an additional noise source in the intensity discrimination task might account for the sensitivity differences measured. Another possibility is that listeners may adopt an optimal decision strategy for the spectral shape discrimination task and a suboptimal decision strategy in the intensity discrimination task. If decision strategy alone governed performance and if the noise in the two tasks was equal, then the condition for which the decision strategy was less optimal would show poorer performance. The current study will evaluate whether hearing-impaired listeners adopt different decision strategies in a spectral shape discrimination task.

Testing hearing-impaired listeners on spectral shape discrimination tasks can be difficult due to the large range of overall level variation needed to discourage intensity discrimination. Kidd and Mason (1992) and Kidd and Dai (1993) proposed a number of experimental manipulations that might be used to reduce the range of level variation needed. One aspect of these techniques is the use of a broadband spectral shape change, in which all components of a standard sound are changed in the signal stimulus. Durlach *et al.* (1986) also suggested that sensitivity would be greatly improved over the traditional spectral shape discrimination task with a single tone by using a signal stimulus generated by increasing the level of half the components of a standard and decreasing the level of the other half (see also Green, 1992). Because thresholds are much lower for this stimulus construction, experimenters can reduce the range of level variation and still be assured that single-channel cues are unreliable. To date, performance of hearing-impaired listeners on these tasks has not been evaluated.

Lentz and Leek (2002) tested the abilities of hearing-impaired listeners to discriminate a spectral shape change involving a single tone added to a standard. Their experiment extended the work of Doherty and Lutfi (1999) who evaluated whether hearing-impaired listeners were able to detect an intensity change in a single component of a multi-tone stimulus. In the presence of level variation, Lentz and Leek found that sensitivity was similar for normal-hearing and hearing-impaired listeners, but large interobserver variability was measured within both groups of listeners. Lentz and Leek also performed a weighting analysis and found that hearing-impaired listeners may have been more likely to adopt suboptimal weighting schemes than normal-hearing listeners. Their conclusion was only tentative, however, as the authors pointed out that differences in internal noise between normal-hearing and hearing-impaired listeners might have been involved in their results.

The current experiment was designed to examine more closely whether hearing-impaired listeners adopt weighting strategies similar to normal-hearing listeners in spectral shape discrimination tasks. The experiment applies a correlation technique to evaluate decision strategies for the discrimination of sounds with different spectral shapes. In contrast to Doherty and Lutfi's work, decision strategies underlying spectral shape discrimination, rather than intensity discrimination, are evaluated. Overall level randomization is used to reduce single-channel intensity cues, and hearing-impaired and normal-hearing listeners are tested at the same stimulus levels. To evaluate the possible contribution of sloping hearing loss on spectral shape discrimination, both normal-hearing and hearing-impaired listeners are tested in two conditions: in one, the standard stimulus is the sum of equal-amplitude tones and, in the other, the standard stimulus is the sum of tones adjusted to compensate for a listener's audiometric thresholds. In both conditions, thresholds and weighting functions are compared between normal-hearing and hearing-impaired listeners to evaluate whether hearing loss leads to a diminished ability to distinguish between sounds with different spectral shapes. For an ideal decision rule, sensitivity and weighting functions should be unaffected by the base stimulus level configuration (see Durlach *et al.*, 1986).

II. METHODS

A. Observer characteristics

Five normal-hearing listeners (NH1–5), ranging in age from 29 to 55 years, and five hearing-impaired listeners (HI1–5), ranging from 60 to 71 years, participated. Hearing losses were moderate and bilateral; the site of lesion was presumed to have cochlear origin based on air- and bone-conduction thresholds and normal results on immitance audiometry. Normal-hearing listeners had pure tone audiometric thresholds within 20 dB of the 1989 ANSI standard between 250 and 4000 Hz. Hearing-impaired listeners were selected to have a pure tone average threshold at 2 and 4 kHz greater than 35 dB HL. For normal-hearing listeners, the right ear always was tested. For the hearing-impaired listeners, the ear with lower audiometric thresholds was the test ear. The audiometric configurations for all listeners are reported in Table I. All observers, except HI5, had prior experience with psychoacoustic experiments, including NH1, who is the first author. One additional hearing-impaired listener enrolled in the study but did not finish. This listener

TABLE I. Audiometric thresholds of test ear for normal-hearing (NH1–NH5) and hearing-impaired listeners (HI1–HI5) (dB HL *re*: ANSI, 1989).

			Frequency (Hz)						
Observer	Age	Test ear	250	500	1000	2000	3000	4000	8000
NH1	29	R	5	10	5	15	15	0	25
NH2	46	R	5	5	5	0	0	0	15
NH3	31	R	0	0	5	5	5	0	0
NH4	42	R	5	5	5	10	10	10	25
NH5	55	R	5	5	5	10	10	10	5
HI1	65	R	55	60	65	70	70	70	60
HI2	68	R	25	30	25	20	30	65	60
HI3	67	L	30	30	50	55	60	55	70
HI4	71	L	10	15	25	35	55	50	50
HI5	60	R	30	40	40	50	50	40	45

was unable to achieve asymptotic performance on all conditions due to the presence of overall level variation.

B. Stimuli

Stimuli were the sum of six tones ranging from 200 to 3000 Hz spaced equidistantly on a logarithmic scale. The ratio between adjacent tones was 1.72. Each time the stimuli were presented, the starting phase of each tone was selected randomly from a uniform distribution ranging from 0 to 2π rad. The stimuli were generated digitally, played by a two-channel 16-bit digital-to-analog converter (DAC, TDT DD1) at a sampling rate of 20 kHz, low-pass filtered at 8.5 kHz (TDT FLT3; approx. 105 dB/oct falloff with a maximum attenuation of 65 dB), and fed into one phone of a TDH-49 headset. The total duration of the stimuli was 300 ms, with 30-ms cosine-squared rise/decay ramps.

Throughout this experiment, three types of stimuli were presented to the listener, reference stimuli, standard stimuli, and signal stimuli. The reference stimulus was always presented in the first of three intervals on every trial as a reminder of the spectral characteristics of the stimuli. On each trial, a comparison was made between the standard and the signal stimuli, presented in random order in the second and third intervals. The characteristics of the different stimulus types are described below.

1. Reference stimuli

The reference stimulus was either the sum of equal-amplitude tones (equal-SPL) or the sum of tones presented at equal sensation levels (equal-SL). For the equal-SPL shaping type, the mean level of each tone was 90 dB SPL. Stimulus components for the equal-SL shaping type were generated to be approximately 20 dB SL. This was accomplished using each listener's audiometric thresholds converted to their SPL values. The frequencies present in the stimuli were not identical to the frequencies tested in the audiometric evaluation, and therefore the adjustments were obtained by linear interpolation of the calibration values specified by the ANSI (1989) standard. To obtain sensation levels at 200 Hz, thresholds were assumed to equal the measured audiometric thresholds at 250 Hz. The actual amount of spectral shaping resulting for the equal-SL stimuli was typically greater for hearing-impaired listeners than for normal-hearing listeners because of the audiogram configurations.

2. Standard stimuli

Standard stimuli were similar to the reference stimuli, but could differ in overall level (due to roving overall amplitudes) or in component levels in some conditions. Stimuli were the sum of tones, having mean component levels of either equal-amplitudes (equal-SPL) or equal sensation levels (equal-SL), depending on the shaping type tested.

3. Signal stimuli

The signal stimulus was generated by increasing the level of three components of the reference stimulus and decreasing the level of the other three components, producing a one-step spectral shape (see Richards and Lentz, 1998). Two signal types were tested. In the ''down–up'' discrimination, the signal had the three low-frequency components decreased in level and the three high-frequency components increased in level. In the ''up–down'' discrimination, the three low-frequency components were increased in level, and the three high-frequency components were decreased in level. Each component was altered by the same amount, ΔL dB, the signal strength.

Two conditions were tested to measure overall sensitivity to changes in spectral shape and spectral weighting functions. For the fixed condition, the reference, standard, and signal stimuli were as described above. For the perturbation condition, the standard and signal were first generated as for the fixed condition, but on every presentation of the stimulus the component levels were further altered by adding zero-mean deviates drawn from a Gaussian distribution with a standard deviation of 1.5 dB to each spectral component. To prevent excessively high and low values, the distribution was truncated so that no deviates exceeded 2.5 standard deviations. The reference stimulus never had component-by-component perturbation added to it. In the perturbation condition, a new selection of component levels was presented to the listener in each interval, so that the same stimulus was not repeated. Schematics of the average long-term power spectra for these three stimulus types in the fixed, equal-SPL condition and perturbation, equal-SPL condition are provided in Fig. 1. All panels show the down–up signal type. The top panels indicate the fixed condition and the bottom panels indicate the perturbation condition.

On every presentation of the stimuli (reference, standard, and signal), the overall levels of the stimuli were varied. The randomization was based on draws from a uniform distribution with a 15-dB range and a 0.1-dB gradation. Stimuli were not allowed to exceed 110 dB SPL. For the equal-SL condition, component levels could be very low due to level variation, component-by-component perturbation, and addition of the signal. Audibility never was a problem, however, as a *post hoc* analysis indicated that the level of any stimulus component fell below 5 dB SL less than 0.01% of the time and fell below 10 dB SL less than 1.5% of the time.

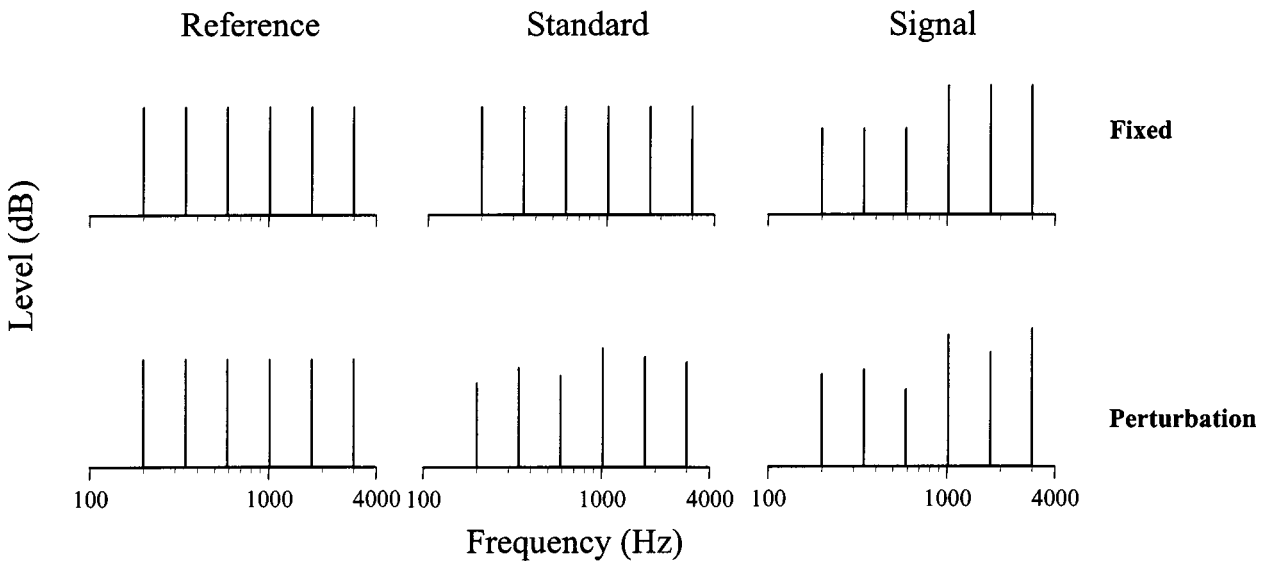


FIG. 1. Sample power spectra for stimuli for an experimental trial for the down–up signal type are shown. The upper and lower panels show fixed and perturbation stimuli, respectively.

C. Procedure

Thresholds reflecting approximately 79% correct detections were estimated using a three-down, one-up procedure (Levitt, 1971). Due to the complexity of the task in the perturbation conditions, a reference/two-alternative forced-choice (2AFC) task was used rather than a two-interval, 2AFC task. Every trial consisted of three intervals. The first interval was the reference stimulus. The second and third intervals contained stimuli that were either signal or standard stimuli chosen randomly. Listeners indicated whether the signal stimulus was presented in interval 2 or interval 3. Note that the listeners could not simply choose the interval of the three that sounded ''different'' because none of the three was exactly the same within a trial.

Two threshold tracks were interleaved within a block of trials, one for the fixed condition and one for the perturbation condition. Whether a trial was fixed or perturbation was chosen at random with equal probability. At the beginning of each track, the mean level of the signal was set to 2–3 dB above each listener's final estimated threshold for the fixed condition. The initial step size of the tracking procedure was 0.5 dB, and after three reversals the step size was reduced to 0.2 or 0.3 dB.[1] Thresholds were estimated with 50 trials in each track. The mean signal level at the last even number of reversal points, excluding the first three, was taken as threshold for each track. Twelve such thresholds were measured for each condition. With the first two counting as practice, ten threshold replicates were averaged to produce the final thresholds reported here. Tracks running to signal levels of 0 dB were encountered by a single listener, NH1, only in the fixed, equal-SPL condition for the up–down signal type. When a 0-dB signal level was encountered, the interval designated to contain the signal stimulus was determined arbitrarily. If the listener correctly guessed that interval as having a signal, the stimulus level remained at 0 dB. No corrections were made to the threshold estimate, which would contain 0-dB values because reversals would occur there. The tracking procedure did not allow the signal level to fall below 0 dB.

Data were collected using a randomized block design. Listeners were tested first on either the equal-SPL or equal-SL shaping type, chosen at random. Within these two shaping types, listeners were tested on the ''up–down'' and the ''down–up'' signals in a random order.

III. RESULTS

A. Sensitivity for normal-hearing and hearing-impaired listeners

Thresholds were measured as ΔL in dB for the blocks of trials with up–down and down–up signal types. An initial repeated measures ANOVA indicated that there were no statistically significant effects involving signal type, so data from these two types of signals were averaged for this analysis. Figure 2 plots thresholds averaged across the five observers in each group as a function of shaping type (equal-SPL or equal-SL) for the fixed and perturbation conditions. Data from normal-hearing listeners are shown as the unfilled bars, and data from hearing-impaired listeners are shown as the hatched bars. Error bars indicate standard errors of the mean across subjects.

A repeated measures ANOVA with one between-subject factor (group membership) and two within-subject factors (shaping type and perturbation/fixed condition) was carried out on the thresholds. Thresholds in the perturbation conditions were higher than thresholds in the fixed conditions by approximately 0.38 dB [$F(1,8)=20.8$; $p<0.005$]. Perturbation increased thresholds by 0.38 and 0.37 dB for the normal-hearing and hearing-impaired listeners, respectively. This increase in threshold under amplitude perturbation has been reported previously, as the uncertainty in the task is greater (Berg and Green, 1990; Kidd *et al.*, 1991; Lentz and Richards, 1998). Spectral shaping also significantly affected sensitivity, with thresholds higher by 0.30 dB for the

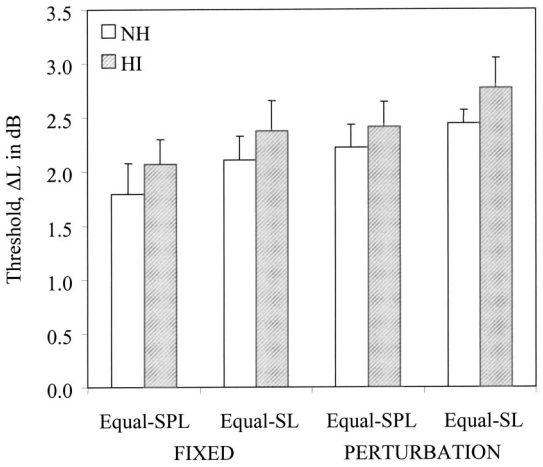


FIG. 2. Averaged thresholds collapsed for down–up and up–down stimuli are plotted for the fixed/equal-SPL, fixed/equal-SL, perturbation/equal-SPL, and perturbation/equal-SL conditions. Unfilled bars denote data for normal-hearing listeners and hatched bars denote data for hearing-impaired listeners. Error bars indicate standard errors of the means.

equal-SL shaping type than for equal-SPL [$F(1,8)=6.3$; $p<0.04$]. Again, similar decreases in performance were obtained in both groups of listeners. Thresholds differed by 0.27 dB for the normal-hearing listeners and by 0.33 dB for the hearing-impaired listeners.

In Fig. 2, the hatched bars are always higher than the unfilled bars, suggesting that sensitivity is slightly poorer for the hearing-impaired listeners. On average, hearing-impaired thresholds are 0.27 dB higher than normal-hearing thresholds. However, the effect of group membership was not statistically significant [$F(1,8)=0.4$; $p=0.53$], nor was this factor involved in any significant interaction. Sensitivity to changes in spectral shape was very similar for normal-hearing and hearing-impaired listeners, and the different experimental manipulations affected thresholds similarly for both groups. This finding suggests some similarities in across-frequency spectral processing for normal-hearing and hearing-impaired listeners.

In general, these thresholds are somewhat higher than reported in other studies. Richards and Lentz (1998) estimated the 79% correct detection threshold in a condition similar to the fixed, equal-SPL condition. In their study, listeners discriminated between an equal-amplitude multi-tone stimulus and a signal stimulus where half of the stimulus components were decreased in level and the other half were increased in level, resembling our up–down signal stimulus. The stimuli used by Richards and Lentz (1998) had four and eight components, whereas the current stimuli had six. In sig *re* stan in dB units [$20\log(\Delta A/A)$; A is the amplitude of a standard component; ΔA is the change in amplitude produced by the signal component] Richards and Lentz measured thresholds of approximately -13 and -17 dB for the four- and eight-component stimuli, respectively. These values are equivalent to ΔL's of 1.75 and 1.15 dB. The present thresholds of 1.79 dB for the normal-hearing listeners and 2.07 dB for the hearing-impaired listeners are only slightly poorer than those of Richards and Lentz (1998) and may be due to less practice given to our listeners. Further, the interleaving of fixed and perturbation tracks, increasing the uncertainty, may have increased thresholds in the fixed conditions.

One goal of the current experiment was to test whether hearing-impaired and normal-hearing listeners have similar sensitivity to changes in the spectral shape of sounds. In order to ensure that listeners were comparing levels across frequency, level variation was included to limit the reliability of a single-channel intensity cue. If listeners based their decisions on the overall level of the stimulus or on the change in level of a single stimulus component, performance would be relatively poor as the level randomization would make changes in stimulus intensity unreliable. For the 15-dB rove used here, an ideal observer that used a single stimulus component to make discriminations would have a threshold of 5.28 dB in the fixed conditions (see Green, 1988). Thresholds obtained for all observers were well below this level, indicating that listeners were using level comparisons of at least two components to distinguish between the standard and the signal.

B. Spectral weighting functions

A weighting analysis describing the relative contribution of each component to a decision statistic was carried out using methods similar to those described by Richards and Zhu (1994) and Lutfi (1995), based on data obtained in the perturbation condition. Stimulus characteristics and responses for the individual trials of the last ten tracks of each subject (500 trials per subject) were used for the analysis which was carried out separately for each frequency component. For each trial, the level difference in dB between interval 2 and interval 3 was determined for each component. The 500 trials then were divided randomly into ten subgroups with 50 discriminations each. For all subgroups, the point biserial correlation coefficient between the level differences and the responses was computed, so that the final correlation coefficients presented here represent the average of ten estimates for each. Eight functions result: two equal-SPL functions (one for up–down and one for down–up signal types) and two equal-SL functions (up–down and down–up) for normal-hearing and hearing-impaired listeners each. Recall that for an ideal decision rule, weighting functions should be equal for equal-SPL and equal-SL shaping types.

Figure 3 indicates weighting functions averaged across listeners to emphasize general trends in the data. Individual weighting functions are plotted in Figs. 4 (NH data) and 5 (HI data). To allow for ease in comparing the effects of experimental variables on the weights, the y axis for the up–down signal type has been inverted in all the figures. Under ideal decision theory, then, the low-frequency weights should fall below the zero-crossing and the high-frequency weights should be above the zero-crossing. In all three figures, filled and unfilled symbols denote data obtained for the equal-SPL and the equal-SL stimulus shapes, respectively. In Fig. 5, the audiometric data for each listener are plotted with the weighting functions. The averaged data will be discussed first, and a discussion of the individual data will follow.

Figure 3 plots data for normal-hearing listeners in the upper panels (triangular symbols), and data for the hearing-

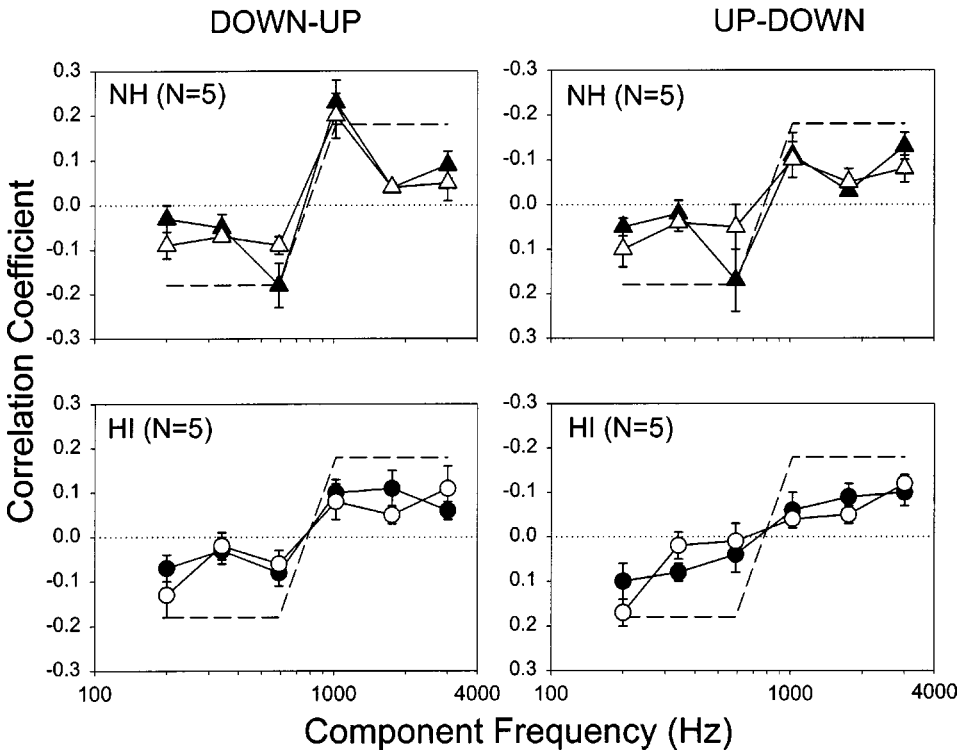


FIG. 3. Correlation coefficients averaged across listeners and signal types are plotted as a function of component frequency. Data from normal-hearing listeners are plotted in the upper panels (triangular symbols), and data from hearing-impaired listeners are plotted in the lower panels (circular symbols). Left and right panels indicate data for down–up and up–down signal types. The ordinate on the right panels has been inverted. In each panel, filled and unfilled symbols denote equal-SPL and equal-SL shaping types, respectively. Error bars indicate standard errors of the mean across five observers. The dashed line indicates the optimal weighting strategy within a constant of proportionality.

impaired listeners in the lower panels (circular symbols). The different signal types are shown in the left and right panels, with down–up in the left panels and up–down in the right panels. The ordinate for the right panels has been inverted. Error bars are the standard errors of the mean across the five listeners. The optimal weighting schemes, within a constant of proportionality, are plotted as the dashed lines in each panel.

Green (1992) derived the optimal weights for the type of stimulus construction used here, where the signal was generated by increasing and decreasing the levels of half of the standard stimulus components by the same amount. An optimal decision strategy would assign equal-magnitude negative weights to the components which had decreased levels from signal to standard and would assign equal-magnitude positive weights to the components which had their levels increased. Positive weights indicate that a listener tended to respond ''signal'' when the level difference of a particular component was high. Negative weights reflect a response of ''signal'' for components having large, negative level differences. Thus, it is expected that the down–up functions would show negative weights for the low-frequencies and positive weights for the high frequencies. The opposite pattern should be seen for the up–down signal type.

For all conditions tested, there is a strong relationship between correlation coefficient and component frequency. The averaged data resemble the optimal weighting strategy somewhat, although some deviations are apparent. Note that any excursions from equal-magnitude correlations may be considered suboptimal. The average correlation coefficients (weights) have signs appropriate to the optimal weighting strategy. As expected, for the down–up signal types, the low-frequency components have negative weights, and the high-frequency components have positive weights. The opposite trend is true for the up–down signal types, where low-frequency components have positive weights, and high-frequency components have negative weights. Neither group of listeners adopted a strategy in which all components had weights of the same magnitude.

In the normal-hearing data (upper panels), all components were not given equal weight, but the weights had the same sign as the optimal predictions. Weights of large magnitude were obtained for the two mid-frequency components and somewhat lower weights for the other components. This finding is most pronounced for the down–up signal type, and two exceptions to this result are noted for the up–down signal type. A large weight is measured for the lowest-frequency component in the up–down, equal-SL condition, and a large weight is assigned to the highest-frequency component for the up–down, equal-SPL condition. To some extent, on average, normal-hearing listeners may have relied more on the mid-frequency components to distinguish the signal from the standard stimulus. Hearing-impaired listeners' weighting strategies also tended to follow the optimal weighting strategy in that the correlation coefficients had the same sign as the optimal prediction. However, they do not show the reliance on the middle-two components that was evident in the normal-hearing data. Their equal-SL data show greater weights at the edges of the spectrum. To a lesser extent the equal-SPL data show this result for the up–down signal type but not for the down–up signal type.

The plots of the individual data allow a more thorough evaluation of the differences in weighting functions that are present across listeners. First, consider the normal-hearing data in Fig. 4. Within a single observer, filled and unfilled symbols somewhat overlap, indicating that the weighting functions are very similar for equal-SPL and equal-SL. Functions also tend to have similar shapes between up–down and

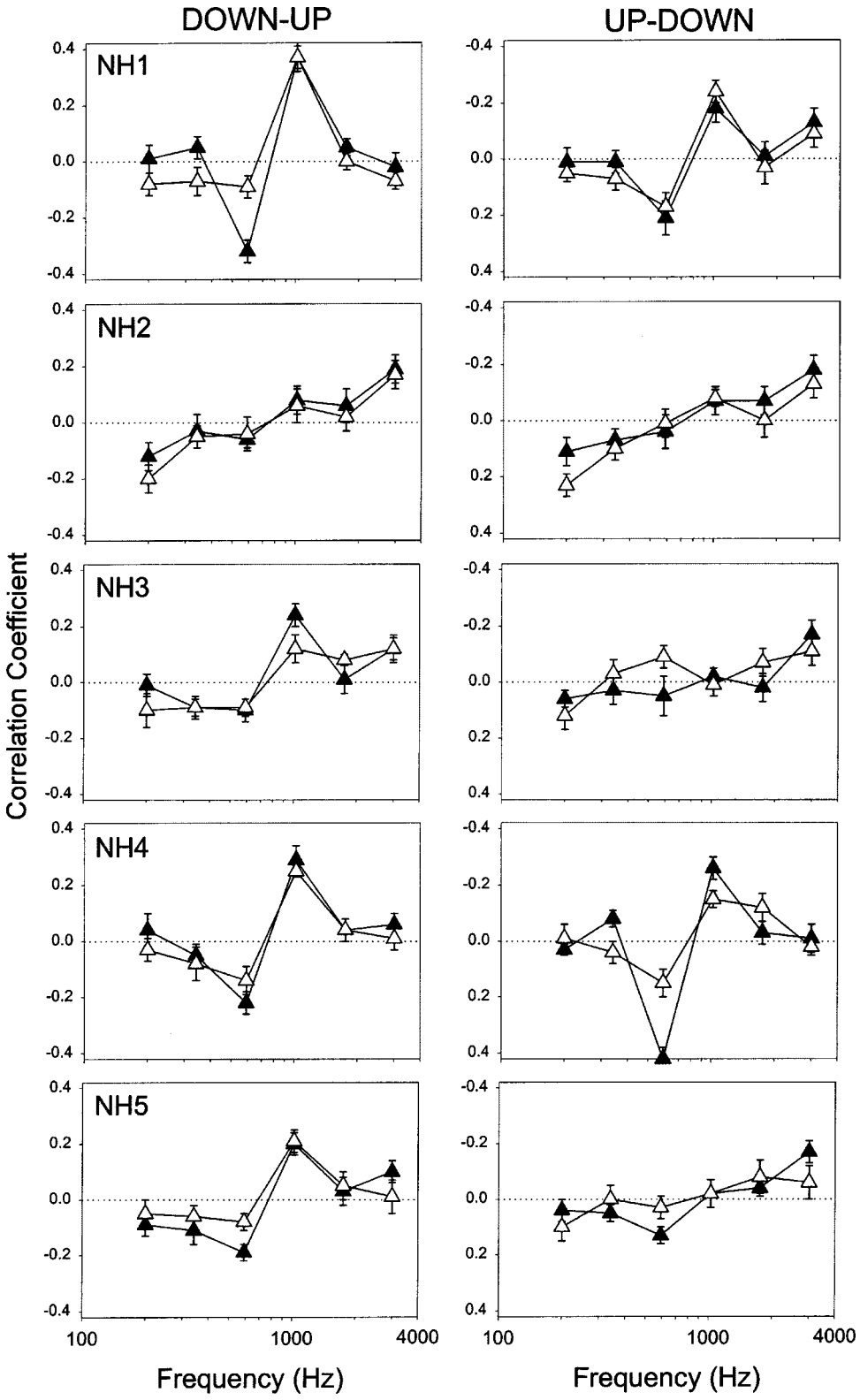


FIG. 4. Individual weighting functions for normal-hearing listeners are presented. Left panels indicate down–up data, and right panels indicate up–down data. Filled and unfilled symbols reflect equal-SPL and equal-SL data. Error bars are the standard errors of the mean across ten correlation coefficient estimates.

down–up signal types. In one example, the data of NH2 indicate a high degree of similarity for all weighting functions. In general, NH2 does not emphasize the two middle components of the stimulus. Rather, NH2 emphasizes the edge components, in contrast with the main trend evident in the averaged data. NH1 and NH4 also exhibit similarity in patterns of weights across frequency. In all conditions, their data indicate that the two central components have the greatest weight. NH3 and NH5 show somewhat more variability than the other three NH listeners across the different stimulus conditions. Variability between the down–up (left panels) and up–down (right panels) signal types is more pronounced than differences between equal-SPL (filled symbols) and equal-SL (unfilled symbols) shaping types. Individually, listeners tend to have very similar weighting functions for equal-SPL and equal-SL, especially in the down–up condi-

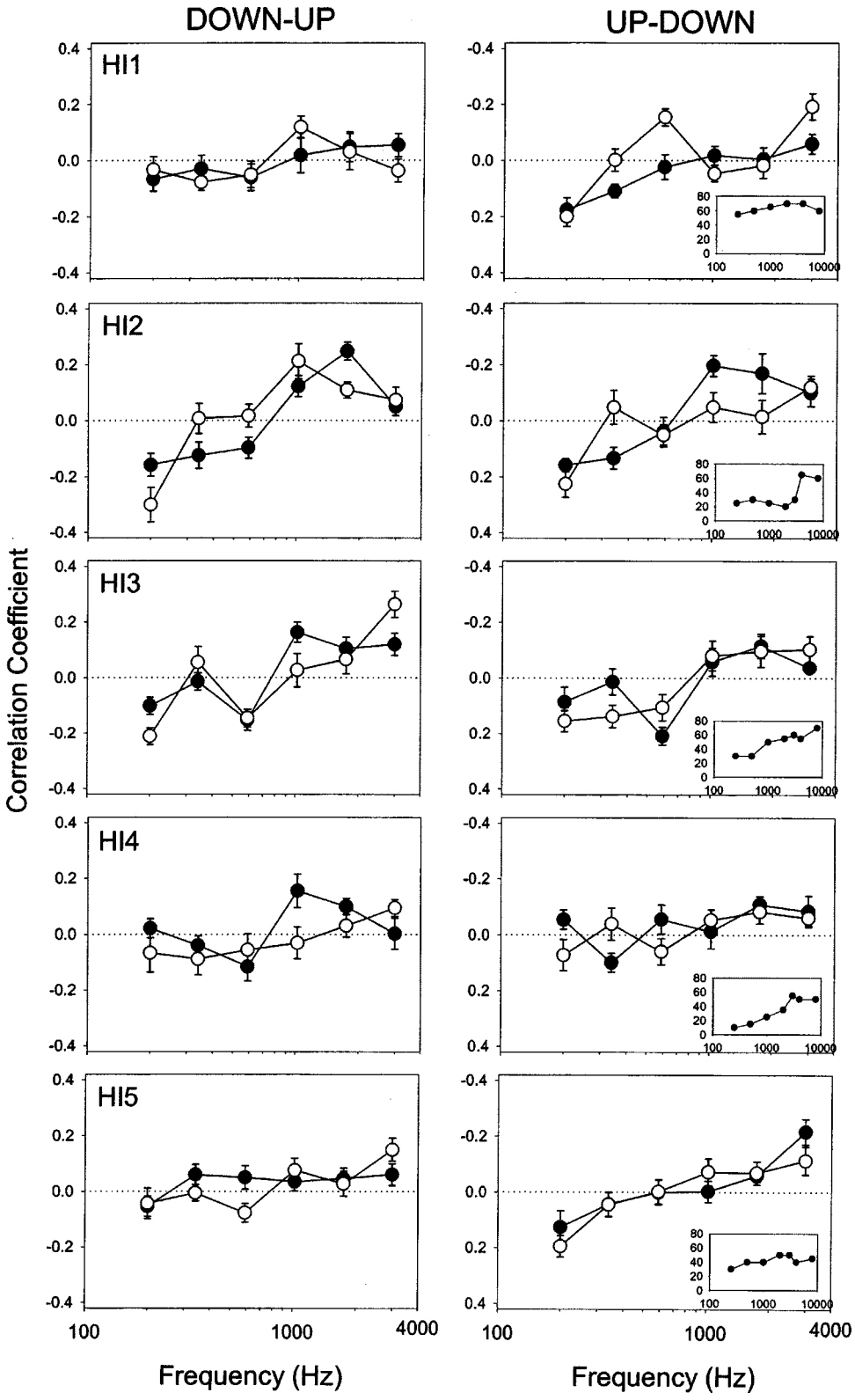


FIG. 5. The format is the same as Fig. 4, only here hearing-impaired data are plotted. Audiograms are plotted in the insets of the right panels. The abscissa on those panels is frequency in Hertz, and the ordinate is dB HL.

tions. Conclusions made from the averaged functions are tempered by the individual data somewhat as not all NH listeners obtained weighting functions consistent with the average.

Data obtained from the hearing-impaired listeners (Fig. 5) show even greater variability across conditions. Depending on the stimulus condition tested, the listener may adopt a weighting strategy that is very different from the strategies adopted in other conditions. A substantial difference between the HI data and those of the NH listeners is that the filled and unfilled symbols generally do not overlap, as was seen in Fig. 4. Large differences between the up–down and down–up conditions also are present. For example, in the down–up conditions HI2 tends to give the greatest weight to the lowest-frequency component in the equal-SL condition and relies the most on the fifth component in the equal-SPL

condition. In the up–down conditions, HI2 still uses the lowest-frequency component to the greatest extent, but the fourth and fifth components have the greatest weights for the equal-SPL condition. In a second example, HI4 utilizes different components to different degrees for all experimental conditions. His equal-SPL, down–up data show reliance on the third and fourth components, whereas his equal-SL, down–up data show reliance on the edge components. His up–down data have a different pattern, with all components having a more equal contribution to the decision, but at times the sign of the weight is not as predicted. For example, the first component for the equal-SPL, up–down data should have a positive sign, but the sign is negative. The components given the most weight by the hearing-impaired listeners vary across listeners and across conditions. A listener who shows an emphasis on the edges in one spectral weighting function (e.g., HI3, equal-SL) does not necessarily rely on the spectral edges for all conditions (HI3, equal-SPL). The substantial variability across listeners and conditions makes further comparisons with the conclusions of Doherty and Lutfi (1996), who found enhanced weights at the ''edges'' of listeners' hearing losses, difficult.

Applying statistics to analyze trends in the data leads to the expectation that the effects of interest will occur in the form of interactions, rather than as main effects. This result is anticipated because the correlation coefficients are, in a sense, ''counterbalanced'' across frequency. That is, each function, when collapsed over frequency, will lead to average correlations of zero. For example, within each weighting function the weights across component frequency involve three negative weights and three positive weights, which will average to zero under ideal decision theory. Functions, when collapsed over signal type, also should lead to average correlations of zero. This expectation only holds if the adopted decision rule is ideal. Weights summing to a value different from zero would suggest a suboptimal decision rule. Any significant main effects that might arise from a statistical analysis would reflect substantial deviations from an ideal decision rule.

In order to determine whether the weighting functions are different depending on group membership and stimulus type (equal-SPL versus equal-SL), the appropriate comparisons are interactions among certain experimental parameters. If listeners adopted optimal weighting strategies, no main effects should be significant. Additionally, any comparison that collapses across component frequency or collapses over signal type should not be significant. The primary interactions of interest include (1) signal type×component frequency to evaluate whether the weighting functions are different for the different signal types (up–down and down–up), (2) group×signal type×component frequency to determine whether the weighting functions differ depending on group membership, (3) stimulus shape×signal type ×component frequency to evaluate whether the weighting functions are different for equal-SPL and equal-SL conditions, and (4) group×stimulus shape×signal type ×component frequency to indicate whether the two stimulus shapes have differential effects on the weighting functions for the two groups of listeners.

A few consequences of comparing correlation coefficients across groups and conditions are worth noting. The magnitude of correlation coefficients is dependent on a number of external and internal parameters. First, the signal level at discrimination threshold impacts the estimate of correlation coefficient (Dai *et al.*, 1996). For these stimuli and decision rule, larger signal levels reduce the magnitude of the correlation coefficient. Internal parameters affecting the magnitude of the correlation coefficient are the variances associated with internal (encoding) noise and central noise (decision noise or response bias). If the internal noise is constant across channels, a large internal noise variance will serve to reduce all correlation coefficients. The presence of the correlated noise source (rove) has a small effect on the magnitude of the correlation coefficient. Assuming that the variance of internal noise does not vary across frequency, the magnitude of all six correlation coefficients of a single function should be affected to the same extent. In addition, the variance of internal noise may vary across frequency, especially for listeners with hearing loss, and would differentially affect correlation coefficients across frequency. Different magnitudes of correlation coefficients may result across observers and conditions. Therefore, significant statistical effects may simply reflect differences associated with the signal level or the presence of noise, rather than a distinctiveness of patterns of weights across frequency.

In the past, researchers have ''normalized'' the correlation coefficients to eliminate any effects of signal level (e.g., Berg and Green, 1990; Doherty and Lutfi, 1996). Another approach has been to introduce catch trials having no signal present and analyzing the weighting functions only on the catch trials (Dai *et al.*, 1996), again eliminating any effects of signal level on correlation coefficients. Here, we have decided against normalizing the correlation coefficients, as we are interested in whether differences in processing are present between normal-hearing and hearing-impaired listeners in spectral shape discrimination tasks. These differences may arise as differences in sensitivity (such as the threshold measures, but also would be observed as lower magnitude of correlation coefficients), differences in pattern of weights across frequency, or differences in internal/central noise sources. Analyzing correlation coefficients may allow a more complete examination of whether any differences in processing are present. Thus, significant effects resulting from the statistical analysis can reflect the impact of all internal and external variables on the estimate of the correlation coefficient.

Interestingly, the ANOVA resulted in two significant main effects: component frequency $[F(5,40)=3.53;\ p<0.01]$ and signal type $[F(1,8)=5.79;\ p<0.05]$. Under ideal decision theory, neither of these main effects should be present, as the correlation coefficients should sum to zero due to the counterbalancing of the experimental design. This result indicates that listeners did not adopt optimal decision strategies. For the effect of component frequency, average correlations tended to be nearest zero for the second and fifth components. The standard deviations of these components also tended to be lower than for the other components. This result indicates that the magnitudes of correlation coefficient

were not identical across the different stimulus conditions (up–down/down–up and equal-SPL/equal-SL). The significant effect of signal type (i.e., up–down versus down–up) suggests that correlations across frequency do not always sum to zero.

The interaction between signal type and component frequency also was significant [$F(5,40)=21.16$; $p<0.001$]. This result was expected, as different patterns of weights should have been adopted for the down–up and up–down signal types (see Fig. 3). The interaction among group membership, signal type, and component frequency did not reach significance [$F(5,40)=2.26$; $p=0.067$], indicating that weighting strategies were similar for both groups of listeners when collapsed across stimulus shape. Two other significant interactions are worth noting. First, the interaction among stimulus shape, signal type, and component frequency reached significance [$F(5,40)=4.88$; $p<0.001$]. This interaction suggests that the weighting functions are different for the equal-SPL and equal-SL shaping types. If listeners were making optimal discriminations, this interaction should not be present, as the spectral shape discrimination process should only depend on the change in level per component and not the base spectral shape of the stimulus (Durlach *et al.*, 1986; Lentz and Richards, 1997). This interaction may be a result of higher thresholds measured for the equal-SL shaping type in which the higher thresholds reduce the magnitude of the correlation coefficients. Finally, a significant four-way interaction (group×stimulus shape×signal type ×component frequency; [$F(5,40)=3.21$; $p<0.02$] suggests that the weighting functions may differ across groups in some complex way. The change in weighting functions imposed by the stimulus shapes and signal types is different for hearing-impaired and normal-hearing listeners.

Because correlation coefficients have been compared, we cannot say that the pattern of weights across frequency is different between the two groups. The results, however, do suggest that some aspect of spectral shape discrimination is affected by hearing loss. For example, hearing loss may slightly decrease spectral shape discrimination sensitivity, thus changing the magnitude of the correlation coefficients. Another possibility is that hearing loss may also be associated with changes in internal or central noise. Whatever the source, differences that were not observed in the threshold measurements arose in the weighting analysis. The patterns of weights across frequency in Fig. 3 suggest that to some degree, hearing-impaired listeners may use different portions of the spectrum to make decisions about simple changes in spectral shape.

IV. DISCUSSION

The goal of the current experiment was to determine whether sensitivity and decision strategies were similar for normal-hearing and hearing-impaired listeners in a profile-analysis task. Despite large differences in the audiometric configurations of the ten listeners in this experiment, hearing-impaired listeners performed as well as normal-hearing listeners when discriminating sounds with different spectral shapes. The increase in thresholds that was measured between the perturbation and fixed conditions and between the equal-SPL and equal-SL shaping type also was similar for both groups of listeners. These two results might indicate that hearing-impaired listeners used a processing scheme that was similar to that used by normal-hearing listeners in this profile-analysis task. The weighting analysis, however, suggested that similarities in thresholds may not necessarily signify similarities in processing because hearing-impaired listeners adopted weights different from normal-hearing listeners, depending on whether equal-SPL or equal-SL was tested.

The result that hearing loss does not lead to elevated thresholds on spectral shape discrimination tasks may not generalize to all stimuli, particularly when components are more closely spaced in frequency than in the stimuli used here. We were interested in evaluating decision strategies under a condition where it was known that spectral smearing, caused by reduced frequency selectivity, would not alter the information that was provided to the central auditory system. Of course, the effects of spectral smearing cannot be eliminated completely, as broadband excitation is produced by passing a stimulus through a complete auditory filter bank. Even stimuli with widely spaced tones are subject to the effects of reduced frequency selectivity. However, the difference in excitation patterns between the signal and standard stimuli used here will be the same regardless of whether the excitation patterns were generated with wider auditory filters or not. Thus, similar amounts of information are presented to the central auditory systems in both healthy and impaired ears.

Past profile-analysis studies have taken a somewhat different approach to evaluate the role that hearing loss has on spectral shape discrimination tasks. Summers and Leek (1994) used stimuli with densely spaced tones but sinusoidally rippled spectral envelopes of varying envelope frequency and suggested that reduced frequency selectivity could account for the differences in sensitivity between normal-hearing and hearing-impaired listeners. Leek *et al.* (1987) also argued that reduced selectivity primarily accounted for poorer performance by hearing-impaired listeners on a synthetic vowel-identification task. These results imply that little functional change occurs in the auditory pathway beyond the cochlea. The current results corroborate these past results to some degree, because we measured no differences in sensitivity between the two groups of listeners. The weighting analysis, however, implicates differences in spectral processing that do not lead to large sensitivity differences between the two groups. The two groups of listeners may be using different portions of the spectrum to make decisions about the sounds presented to them. Normal-hearing listeners possibly used the central portions of the spectrum, whereas hearing-impaired listeners may have been more likely to use the edges of the spectrum. Spectral smearing, caused by reduced frequency selectivity in impaired ears, may cause these listeners to direct their attention away from the central portions of the spectrum and toward the edges of the spectrum, even though the ''information'' (or the change in level between signal and standard stimuli) is the same for both groups. Other factors, such as differences in internal and/or central noise, may also be present between

the two groups of listeners. Differences in performance that might arise due to different weights would be difficult to measure because all listeners adopted suboptimal weighting strategies. Similar sensitivity across subjects in the presence of different weighting strategies suggests that other factors, such as additional noise or stimulus uncertainty, contribute largely to performance.

The variability present in the weighting functions obtained from the hearing-impaired listeners has implications for the interpretation of the threshold data. Theoretically, weights should have been the same for all conditions, with the exception that weights for the down–up signal be the inverse of weights for up–down. The normal-hearing listeners showed some variability across the conditions, but less than the hearing-impaired listeners. The result that each hearing-impaired listener tended to adopt a different weighting strategy across stimulus conditions suggests that cochlear hearing loss leads to difficulty in making comparisons of spectral shape. The variability in the measurements may reflect noise associated with the hearing loss, such as difficulty encoding a reference stimulus, difficulty using a reference stimulus, or greater variability in encoding the stimulus being presented.

A second goal of the current study was to determine whether spectral shaping of the standard stimulus, compensating for the configuration of hearing loss, would lead to differences in performance. At the outset of the study, we had thought that relative amplification of the high-frequency components might lead to better thresholds for the hearing-impaired listeners who had sloping high-frequency losses. We also expected that sensitivity would not be affected by spectral shaping for the normal-hearing listeners. Both of these hypotheses were incorrect, however, as thresholds measured for the equal-SL shaping type always were higher than thresholds for equal-SPL. One possible reason for this finding in the normal-hearing listeners is that overall presentation level may affect performance. For these listeners, the overall presentation level for equal-SL was much lower than the presentation level for equal-SPL. Hearing-impaired listeners also were tested at different overall levels between equal-SL and equal-SPL stimuli, but the overall level of the equal-SL stimuli varied substantially across listeners. A study by Mason *et al.* (1984) casts doubt on the possibility that performance is affected by stimulus level. Mason *et al.* measured profile-analysis thresholds for a level increment to a single tone of an equal-amplitude standard presented at 30, 40, 50, 60, and 70 dB SPL per component. They found only small, unsystematic differences in performance at the different stimulus levels. The reason for the elevated thresholds in the current study is unclear, but other profile-analysis studies have found that level comparisons across frequency are easier when the spectral regions to be compared are relatively close in level at their input. Kidd *et al.* (1986) presented data in which listeners detected a level increment added to a standard which had a ''jagged'' spectral shape. The levels of the components were chosen randomly, but unlike the perturbation condition, the levels were fixed across all stimulus presentations. Kidd *et al.* (1986) showed that thresholds increased as the standard deviation of the fixed-amplitude change was increased (see also Lentz and Richards, 1998). These results emphasize that the auditory system does not perform as an optimal detector. If listeners adopted the optimal weighting strategy, then sensitivity should only depend on the change in level per component (e.g., Durlach *et al.*, 1986) and be independent of any spectral variations in the standard. In cases where the standard has large level changes across the spectrum, listeners may have difficulty encoding a long-term reference stimulus for spectral comparisons. Hearing-impaired listeners showed the same effect as normal-hearing listeners, in that thresholds were higher for the equal-SL shaping type. Shaping the stimulus to accommodate their hearing losses did not provide additional assistance to these listeners on this task. Their thresholds probably are elevated for the same reason as the normal-hearing listeners, i.e., large changes in level across the stimulus spectrum lead to difficulties in comparing spectral shapes.

The weighting analysis was intended to support conclusions based on the threshold analysis and also was used to identify differences in decision strategies adopted by normal-hearing and hearing-impaired listeners in spectral shape discrimination. One primary aspect of the weighting data was that neither group of listeners adopted weighting strategies that were optimal. In many profile analysis studies, listeners adopted the optimal weighting strategy (Berg and Green, 1990; Dai *et al.*, 1996). These studies used a spectral change in which a single component, near the center of the spectrum, was altered between standard and signal stimuli. Other studies have used more complex spectral changes and found that listeners do not always adopt the optimal weighting strategy. Berg and Green (1992) measured weights in a condition where listeners discriminated between an eight-tone equal-amplitude stimulus and a signal stimulus with a ''rippled'' spectral shape generated by imposing a level change consisting of one cycle of a sine wave across the eight components. Here, listeners did not consistently demonstrate the optimal weighting strategy. When Berg and Green (1992) tested a signal stimulus that was generated using two cycles of a sine wave, listeners had even greater difficulty adopting optimal weighting strategies. Matiasek and Richards (1999) used a stimulus set in which every component of an equal-amplitude, multi-tone standard was changed by alternately increasing and decreasing every other stimulus component. In almost all instances, their listeners did not adopt optimal weighting strategies.

One question to ask, then, is whether the weighting functions of hearing-impaired listeners are more or less different from the optimal weighting strategy than normal-hearing listeners and whether this difference is more pronounced in either the equal-SPL or the equal-SL shaping type (i.e., to explore the four-way interaction found in the weighting data). This analysis also allows a test of whether the differences in sensitivity between equal-SPL and equal-SL are due to adoption of less optimal weighting strategies for the equal-SL shaping type. To determine how optimal a particular weighting scheme was, we used an efficiency analysis as proposed by Berg (1990). The efficiency analysis was applied to normalized data, in which each func-

TABLE II. Weighting efficiency values, η_{wgt}, for normal-hearing (NH) and hearing-impaired (HI) observers for the different shaping types (equal-SPL and equal-SL) and different signal types (down–up and up–down).

Observer no.	NH				HI			
	Equal-SPL		Equal-SL		Equal-SPL		Equal-SL	
	down–up	up–down	down–up	up–down	down–up	up–down	down–up	up–down
1	0.30	0.54	0.30	0.56	0.88	0.54	0.48	0.05
2	0.76	0.79	0.62	0.59	0.83	0.87	0.49	0.39
3	0.61	0.44	0.97	0.14	0.84	0.65	0.50	0.95
4	0.45	0.29	0.55	0.50	0.52	0.18	0.59	0.59
5	0.81	0.66	0.60	0.68	0.08	0.49	0.54	0.64
AVG	**0.59**	**0.54**	**0.61**	**0.50**	**0.63**	**0.55**	**0.54**	**0.53**
(std error)	(0.09)	(0.09)	(0.11)	(0.09)	(0.15)	(0.11)	(0.03)	(0.15)

tion was scaled so that the correlation coefficient having the largest magnitude was assigned a value of 1.0 or −1.0, depending on its original sign. Efficiency, $\eta_{wgt}=(d'_{wgt}/d'_{opt})^2$, was calculated using the optimal and obtained weights. The equation for d' is $\Sigma a_i \Delta\mu_i / \sqrt{\Sigma a_i^2 \sigma_i^2}$, where a_i are the weights for each channel, i; $\Delta\mu_i$ are the signal levels applied to each component; and σ_i^2 is the variance of an internal noise distribution (Berg, 1990). For the efficiency analysis, σ_i was assumed to be equal for all i. The efficiency values are listed in Table II. An efficiency of 1.0 would indicate that the listener adopted the optimal weighting strategy, and the lower the efficiency, the greater the deviation from the optimal weighting strategy. A repeated-measures ANOVA indicates no significant effects of the experimental variables on efficiency.

Average efficiencies are very similar for the equal-SL ($\eta_{wgt}=0.54$) and equal-SPL ($\eta_{wgt}=0.58$) shaping type. This result suggests that listeners were not more likely to adopt suboptimal weighting schemes for equal-SL. The result that efficiencies do not differ between equal-SPL and equal-SL suggests that the sensitivity data in Fig. 2 are not due to listeners adopting suboptimal weighting schemes for the equal-SL shaping type. A more likely interpretation, then, is that thresholds might be elevated due to more uncertainty in the task or more difficulty in encoding the reference stimulus. This result also indicates that weighting strategy alone does not govern performance on this task. A negative correlation between thresholds and efficiency ($\rho=-0.28$) indicates that as thresholds increase, efficiency tends to decrease. This correlation suggests that deviation from optimal weighting strategies can lead to increased thresholds. On average, efficiencies were similar for normal-hearing and hearing-impaired listeners ($\eta_{wgt}=0.56$ for both groups), but large variability in efficiencies was present across listeners. Normal-hearing listeners have efficiencies ranging from 0.14 to 0.97, and hearing-impaired listeners' range from 0.05 to 0.95.

Overall, the efficiency results suggest that whatever differences are present between these two groups, hearing-impaired listeners do not appear to adopt weighting strategies that are less optimal than normal-hearing listeners. Doherty and Lutfi (1996) also showed that weighting efficiencies were not different between the two groups of listeners. Yet, their data revealed differences in the weighting functions between the two groups, and the authors suggested that hearing-impaired listeners tended to give the greatest weight to frequency components near the edge of their hearing loss. Our results agree with the findings that hearing-impaired listeners may not adopt the same weighting strategy as normal-hearing listeners, but our data cannot speak to the suggestion that listeners rely on frequency components near the edges of their hearing loss. A large difference between the study by Doherty and Lutfi and the current study is that the listeners tested by Doherty and Lutfi all had a substantial edge to their hearing losses. In this study, only listeners HI2–4 had edges to their hearing losses, and the edges were not as pronounced as in the listeners of Doherty and Lutfi. For this reason, we cannot evaluate whether listeners rely more on the edges of their hearing loss in this spectral shape discrimination task.

V. SUMMARY AND CONCLUSIONS

Normal-hearing and hearing-impaired listeners showed similar sensitivity on a spectral shape discrimination task. The stimuli were constructed so that limitations in audibility were not factors in this experiment and so that the "information" between signal and standard stimuli was similar for the two groups of listeners. Differences were measured in the weighting functions between normal-hearing and hearing-impaired listeners, but these differences did not lead to poorer sensitivity for one group over the other. In fact, both normal-hearing and hearing-impaired listeners appeared to adopt suboptimal weighting schemes, and some differences were measured in the weighting functions between the two groups. These differences were manifested in the way in which the weighting functions changed, depending on whether the tones in the stimuli were presented at equal sensation levels or equal sound pressure levels.

Compensating for a listener's hearing loss by using a stimulus made of tones presented at equal sensation levels led to more difficulty in detecting a change in spectral shape for both groups of listeners. Contrary to our expectation, the amplification of components in the region of a listener's hearing loss with respect to other components did not improve performance. It may be that listeners are better at discriminating spectral shape changes when the sound pressure levels of the tones are similar and that listeners had more difficulty forming a long-term standard for the equal-SL shaping type. If poorer performance with relative amplifica-

tion that is not uniform across frequency generalizes to more complex stimuli, there may be broad implications for characteristics of amplification provided to hearing-impaired listeners. In some instances, normal-hearing listeners used the central portion of the spectrum to make discriminations, which is the frequency location where the largest change in level occurred between adjacent stimulus components. Hearing-impaired listeners, however, did not show this emphasis on the central components and at times may have used the edges of the spectrum to a greater degree. One possible explanation for this result is that the spectral smearing resulting from hearing loss may lead listeners to divert their attention to the edges of the spectrum. However, listeners with hearing loss showed substantial variability across conditions, and no one listener used the edges of the spectrum for all conditions. The variability suggests that hearing loss leads to difficulty in the encoding of stimuli and storing/recalling the references that are needed to discriminate between sounds with different spectral shapes.

These results, in conjunction with those of Doherty and Lutfi (1996, 1999), suggest that cochlear damage can lead to differences in the manner in which information is compared and integrated across frequency. Large differences in optimality of the weighting functions were measured across listeners, however, suggesting that performance on these tasks is not primarily governed by the weighting strategy. The adoption of different weighting strategies by the two groups of listeners may not be of primary importance in simple tasks such as these. It remains to be seen, however, whether alterations in these more central processes lead to differences in the way more complex sounds, such as speech and music, are perceived by listeners with cochlear hearing loss.

ACKNOWLEDGMENTS

This work was supported by Grant No. DC 00626 from the National Institute on Deafness and Communication Disorders and was carried out under Department of Clinical Investigation Work Unit #2587, Walter Reed Army Medical Center. The authors thank Laura Dreisbach and Lina Kubli for assistance during data collection. We are grateful to Michelle Molis and Ken Grant for their valuable comments and interesting discussions on the data presented. All subjects participated voluntarily and provided written, informed consent. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

[1]In the computer-controlled algorithm, the smaller step size was intended to be 0.25 dB. However, the attenuators used in the experiment only have a resolution of 0.1 dB, and therefore could not represent 0.25-dB changes. The values imposed by the attenuator were rounded as the signal level (in ΔL units) was changed.

ANSI (**1989**). ANSI S3.6-1989, ''Specifications for audiometers'' (American National Standards Institute, New York).

Berg, B. G. (**1989**). ''Analysis of weights in multiple observation tasks,'' J. Acoust. Soc. Am. **86**, 1743–1746.

Berg, B. G. (**1990**). ''Observer efficiency and weights in a multiple observation task,'' J. Acoust. Soc. Am. **88**, 149–158.

Berg, B. G., and Green, D. M. (**1990**). ''Spectral weights in profile listening,'' J. Acoust. Soc. Am. **88**, 758–766.

Berg, B. G., and Green, D. M. (**1992**). ''Discrimination of Complex Spectra: Spectral Weights and Performance Efficiency,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, London), pp. 373–379.

Bernstein, L. R., and Green, D. M. (**1987**). ''The profile-analysis bandwidth,'' J. Acoust. Soc. Am. **81**, 1888–1895.

Dai, H., Nguyen, Q., and Green, D. M. (**1996**). ''Decision rules of listeners in spectral-shape discrimination with or without signal-frequency uncertainty,'' J. Acoust. Soc. Am. **99**, 2298–2306.

Doherty, K. A., and Lutfi, R. A. (**1996**). ''Spectral weights for overall level discrimination in listeners with sensorineural hearing loss,'' J. Acoust. Soc. Am. **99**, 1053–1058.

Doherty, K. A., and Lutfi, R. A. (**1999**). ''Level discrimination of single tones in a multitone complex by normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **105**, 1831–1840.

Durlach, N. I., Braida, L. D., and Ito, Y. (**1986**). ''Towards a model for discrimination of broadband signals,'' J. Acoust. Soc. Am. **80**, 63–72.

Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford U. P., New York).

Green, D. M. (**1992**). ''The number of components in profile analysis tasks,'' J. Acoust. Soc. Am. **91**, 1616–1623.

Green, D. M., and Mason, C. R. (**1985**). ''Auditory profile analysis: Frequency, phase and Wèber's Law,'' J. Acoust. Soc. Am. **77**, 1155–1161.

Green, D. M., Kidd, G., Jr., and Picardi, M. C. (**1983**). ''Successive versus simultaneous comparison in auditory intensity discrimination,'' J. Acoust. Soc. Am. **73**, 639–643.

Kidd, G., Jr., and Dai, H. (**1993**). ''A composite randomization procedure for measuring spectral shape discrimination,'' J. Acoust. Soc. Am. **94**, 1275–1280.

Kidd, G., Jr., and Mason, C. R. (**1992**). ''A new technique for measuring spectral shape discrimination,'' J. Acoust. Soc. Am. **91**, 2855–2864.

Kidd, G., Jr., Mason, C. R., and Green, D. M. (**1986**). ''Auditory profile analysis of irregular sound spectra,'' J. Acoust. Soc. Am. **79**, 1045–1053.

Kidd, G., Jr., Mason, C. R., Uchanski, R. M., Brantley, M. A., and Shah, P. (**1991**). ''Evaluation of simple models of auditory profile analysis using random reference spectra,'' J. Acoust. Soc. Am. **90**, 1340–1354.

Leek, M. R., Dorman, M. F., and Summerfield, Q. (**1987**). ''Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **81**, 148–154.

Lentz, J. J., and Leek, M. R. (**2002**). ''Decision strategies of hearing-impaired listeners in spectral shape discrimination,'' J. Acoust. Soc. Am. **111**, 1389–1398.

Lentz, J. J., and Richards, V. M. (**1997**). ''Sensitivity to changes in overall level and spectral shape: An evaluation of a channel model,'' J. Acoust. Soc. Am. **101**, 3625–3635.

Lentz, J. J., and Richards, V. M. (**1998**). ''The effects of amplitude perturbation and increasing numbers of components in profile analysis,'' J. Acoust. Soc. Am. **103**, 535–541.

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

Lutfi, R. A. (**1995**). ''Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks,'' J. Acoust. Soc. Am. **97**, 1333–1334.

Mason, C. R., Kidd, G., Jr., Hanna, T. E., and Green, D. M. (**1984**). ''Profile analysis and level variation,'' Hear. Res. **13**, 269–275.

Matiasek, M. M., and Richards, V. M. (**1999**). ''Relative weights for three different psychophysical tasks,'' J. Acoust. Soc. Am. **106**, 2209–2210(A).

Richards, V. M., and Lentz, J. J. (**1998**). ''Sensitivity to changes in level and envelope patterns across frequency,'' J. Acoust. Soc. Am. **104**, 3019–3029.

Richards, V. M., and Zhu, S. (**1994**). ''Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients,'' J. Acoust. Soc. Am. **95**, 423–434.

Summers, V., and Leek, M. R. (**1994**). ''The internal representation of spectral contrast in hearing-impaired listeners,'' J. Acoust. Soc. Am. **95**, 3518–3528.

Speech perception, localization, and lateralization with bilateral cochlear implants

Richard J. M. van Hoesel[a)]
CRC for Cochlear Implant and Hearing Aid Innovation, Melbourne, Victoria, Australia

Richard S. Tyler
The Departments of Otolaryngology—Head and Neck Surgery and Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa

(Received 6 June 2002; revised 21 November 2002; accepted 22 November 2002)

Five bilateral cochlear implant users were tested for their localization abilities and speech understanding in noise, for both monaural and binaural listening conditions. They also participated in lateralization tasks to assess the impact of variations in interaural time delays (ITDs) and interaural level differences (ILDs) for electrical pulse trains under direct computer control. The localization task used pink noise bursts presented from an eight-loudspeaker array spanning an arc of approximately 108° in front of the listeners at ear level (0-degree elevation). Subjects showed large benefits from bilateral device use compared to either side alone. Typical root-mean-square (rms) averaged errors across all eight loudspeakers in the array were about 10° for bilateral device use and ranged from 20° to 60° using either ear alone. Speech reception thresholds (SRTs) were measured for sentences presented from directly in front of the listeners (0°) in spectrally matching speech-weighted noise at either 0°, +90° or −90° for four subjects out of five tested who could perform the task. For noise to either side, bilateral device use showed a substantial benefit over unilateral device use when noise was ipsilateral to the unilateral device. This was primarily because of monaural head-shadow effects, which resulted in robust SRT improvements ($P<0.001$) of about 4 to 5 dB when ipsilateral and contralateral noise positions were compared. The additional benefit of using both ears compared to the shadowed ear (i.e., binaural unmasking) was only 1 or 2 dB and less robust ($P=0.04$). Results from the lateralization studies showed consistently good sensitivity to ILDs; better than the smallest level adjustment available in the implants (0.17 dB) for some subjects. Sensitivity to ITDs was moderate on the other hand, typically of the order of 100 μs. ITD sensitivity deteriorated rapidly when stimulation rates for unmodulated pulse-trains increased above a few hundred Hz but at 800 pps showed sensitivity comparable to 50-pps pulse-trains when a 50-Hz modulation was applied. In our opinion, these results clearly demonstrate important benefits are available from bilateral implantation, both for localizing sounds (in quiet) and for listening in noise when signal and noise sources are spatially separated. The data do indicate, however, that effects of interaural timing cues are weaker than those from interaural level cues and according to our psychophysical findings rely on the availability of low-rate information below a few hundred Hz. [DOI: 10.1121/1.1539520]

PACS numbers: 43.66.Pn, 43.66.Ts [LRB]

I. INTRODUCTION

For normal-hearing listeners, speech is often easier to understand in noise and sounds are easier to localize using two ears rather than one (e.g., Carhart, 1965; Dirks and Wilson, 1969; Hirsch, 1950; Licklider, 1948; MacKeith and Coles, 1971; Peissig and Kollmeier, 1997; Hawley *et al.*, 1999). When speech and noise are spatially separated, different signal-to-noise ratios (S/N) result at the two ears because of the monaural head-shadow for each ear. The listener can take advantage of this by attending to the ear with the better S/N. In addition, binaural unmasking benefits (sometimes referred to as squelch) are available from the comparison of information at both ears. It is also the use of information from both ears, in particular the comparison of interaural level differences (ILDs) and interaural time-delays (ITDs), that greatly facilitates localization in the horizontal plane (Rayleigh, 1907; Stevens and Newman, 1936; Searle *et al.*, 1976; Hausler *et al.*, 1983; Bronkhorst and Plomp, 1988, 1989; Wightman and Kistler, 1992). Work in the early 1990s by van Hoesel *et al.* (1993) and van Hoesel and Clark (1995) with two bilateral cochlear implant (CI) users showed that these subjects could readily fuse information from two devices. Although these two subjects showed poor sensitivity to ITDs, with just noticeable differences (jnd's) of the order of 1 ms, they showed strong effects of level variations across the ears and experienced loudness summation effects comparable to normal hearing (van Hoesel and Clark, 1997). Compared to those subjects, a case study presented by Lawson *et al.* (1998) showed considerably better ITD sensitivity around a few hundred μs and a recent progress-report from the same laboratory shows ITD effects with cues as small as 50 μs for several bilateral CI users.[1] A bilateral

a)Electronic mail: r.vanh@medoto.unimelb.edu.au

CI-user studied by Long (2000) also showed consistent lateralization shifts resulting from signals with ITDs of 300 μs. Speech tests with one of the early bilaterally implanted subjects studied by van Hoesel and Clark (1999) showed that the subject was able to take advantage of the head-shadow by attending to the ear with the better S/N for spatially separated speech and noise. Similarly, in a more recent case study by van Hoesel *et al.* (2002), a clear head-shadow advantage was demonstrated but little evidence of binaural unmasking was observed. However, the authors could not rule out the possibility that this was because of ceiling effects resulting from the fixed S/N levels used in their speech tests. The same study also showed a large benefit in identifying sound-direction with two cochlear implants compared to just one when tested with 11 loudspeakers spanning 180° in the frontal horizontal arc in an anechoic room. That benefit was relatively unaffected by the variations in coding parameters the authors tested with the subject's sound-processor. The variations predominantly compared higher electrical stimulation rates and filtering methods that preserved faster envelope fluctuations than those used in the subject's clinical device, to test the hypothesis that better preservation of envelope-ITDs would offer binaural advantages with the electrical stimulation.

Further speech data and some limited localization data with bilateral implant users have been reported recently by Tyler *et al.* (2002), Ganz *et al.* (2002), and Müller and Helms (2002). In the study by Tyler *et al.*, data from nine subjects were collected three months after bilateral implantation. Speech-intelligibility was tested both in quiet and in noise and the ability to discriminate a signal arriving from 45° to the left from one arriving from the right was assessed. When the noise was separated spatially from the speech signals, the subjects showed a significant advantage from the head-shadow but only a few of them displayed evidence of an additional binaural benefit. Without interfering noise, a few subjects also showed improved word-scores binaurally compared to using either ear alone. Subjects were allowed to adjust volume settings for subjective preference when switching between left, right, or binaural listening. This procedure allows an assessment of how performance might be affected if the mapping in the processor is not adjusted specifically to allow for monaural or binaural use and a volume or sensitivity control is used instead. However, it may also introduce uncontrolled loudness variations and potential S/N shifts in the comparison between one and two ears if the acoustic-to-electrical loudness-mapping gradient is not preserved (see van Hoesel *et al.*, 2002). Most subjects in the study also showed substantially better discrimination of left and right signals at 45° using two ears compared to just one, although some subjects could perform the task monaurally as well. The study was followed up one year after implantation with measurements from the same subjects, with one additional subject included and using the same methods (Gantz *et al.*, 2002). The results of the follow-up study showed comparable results. Müller and Helms (2002) have also reported speech data from ten bilateral implant users. In their study, subjects were specifically not allowed to change the levels on the processors to compensate for loudness summation effects when switching between monaural and binaural listening. The results for sentences (at 0°) in noise (at $\pm$90°) and a S/N of 10 dB again show a significant head-shadow benefit as well as an additional benefit of using both ears over the shadowed one. Performance for bilateral implants for monosyllables in quiet also showed improvement compared to the better ear alone.

Unfortunately, in the speech studies described above, benefits were recorded as percentage point increases for the particular test materials and comparison of results across studies is not straightforward. Furthermore, fixed signal-to-noise ratios were used which, particularly for spatially separated sound-sources, can easily be dominated by floor or ceiling effects. The Tyler *et al.* (2002), Ganz *et al.* (2002), and Müller and Helms (2002) studies also make no mention of any acclimatization periods for new listening conditions when swapping between binaural and monaural conditions and it is possible that this contributes to the bilateral condition having an unfair advantage because it is the condition that subjects use in everyday life.[2] This may lead to overestimating binaural benefits in noise as well as in quiet. Binaural loudness summation may also have affected comparisons between monaural and binaural performance in different ways in these studies. Although it seems likely that, as also mentioned in Müller and Helms (2002), binaural loudness summation will affect speech in quiet more than in noise, it is not clear to us that there will be no effect in noise. In a linear system, it is true that assuming equal impact of summation on the signal and noise the S/N would remain unchanged, but this is not necessarily so in an implant system where there is a considerably more limited input dynamic range compared to normal hearing. Note also that all the experiments to date have been conducted with standard clinical sound-processing strategies, which were not developed with bilateral application in mind and may not be well suited to preserving binaural cues.

The studies presented in this manuscript aimed to examine speech-in-noise and localization benefits available to bilateral implant users by testing five subjects with a broad range of listening tasks. A new sound-processing strategy that aimed to better preserve fine-structure ITDs was used in several of these tests. An improved assessment of head-shadow and binaural unmasking contributions for speech in noise was made by using an adaptive signal-to-noise ratio (S/N) method to avoid ceiling and floor effects. An attempt was made to take the effects of binaural loudness summation and acclimatization to new listening conditions into consideration. Our main motivation for eliminating loudness contributions that could potentially affect intelligibility was to try to better assess the binaural benefit that results from comparison of the signals at the two ears *per se*. Another reason was to compare what might be clinically relevant situations; because an implant system should be optimized for users' comfort, we would expect that similar loudness should be targeted irrespective of whether they are fitted unilaterally or bilaterally. Separate maps for binaural and monaural listening were therefore used, taking care to preserve the gradient of the acoustic-to-electrical loudness-mapping function for the two cases. Similar localization tests were used as those

TABLE I. Subject details for participants in these studies. ''Age'' is age at time of first electrical stimulation, which was at least 12 months prior to the studies described here. ''Years of deafness'' is years from severe-profound hearing loss in each ear until first stimulation with cochlear implants, as assessed retrospectively by subjects.

Subject	Aetiology	Age (years)	Years of deafness	
			L	R
A	Meniere's	59	1	2
B	Meniere's	69	25	8
C	Autoimmune	36	7	18
D	Unknown	59	5	5
E	Meniere's	71	0.5	0.5

described in van Hoesel *et al.* (2002), except now an eight-loudspeaker array spanning 108° was used and the analysis methods were expanded to include examination of results for individual loudspeaker positions. Complementary psychophysical studies were conducted to help separate sound processing from perceptual limitations with electrical stimulation via the implant systems used by these subjects. For those studies stimuli were generated by controlling stimulation currents at the electrodes in each ear under direct computer control rather than by presenting signals via sound-coding strategies.

As a cautionary note, in the normal-hearing literature the term ''localization'' is often reserved for experiments in which subjects experience extracranial sound-images. Unless stimuli are filtered by head-related transfer functions, headphone listening tasks usually result in percepts that are intracranial instead, and usually the term lateralization is used to describe the ''sidedness'' of the percept. For convenience, we have adopted the term ''localization'' when discussing implant users' ability to identify sound-direction when actual sound sources were external and acoustic signals were picked up by the microphones of the subjects' sound-processors. In our use of the term in this paper, we do not imply that the subjects necessarily experienced extra-cranial sound-images. This was not explicitly tested. We have used the term ''lateralization'' for those experiments in which electrical stimuli were under direct computer control and the listener was asked to make judgments about the ''sidedness'' of resulting sound-images, which were presumably intracranial.

II. SUBJECTS, SOUND-PROCESSING, AND STIMULI

Five bilateral users of the Nucleus CI-24M implant system who were implanted at the University of Iowa participated in the studies. They had received implants in both ears in a single surgery and had used bilateral sound processors from the outset. The implant system uses pulsatile, nonsimultaneous biphasic current pulses on up to 22 platinum bands, spaced at 0.75-mm intervals along a silicone carrier placed inside the Scala Tympani. All five participants demonstrated a difference between ears in either length of time of deafness before implantation, preimplantation pure-tone thresholds, or both (Table I). To take part in the study, patients visited the University Hospital in Iowa City on two occasions. On the first visit, which typically lasted 2 to 3 days, localization and speech-in-noise performance were tested using the clinical sound-processors the subjects had been using for at least 12 months prior to these tests. With these processors, no attempt had been made to compensate for any insertion offsets between the arrays in each ear and the manufacturer's default frequency-to-electrode maps were used. Some subjects used two Cochlear Esprit behind-the-ear (BTE) processors, which include directional microphones, and others used body-worn Cochlear Sprint processors with Cochlear HS8 microphones behind each ear. However, the differences between the two are not important with regard to the issues discussed in this paper. The same microphone elements (Knowles EL 7189) are used in both and the ports are spaced such that similar directional characteristics result. During the first visit, subjects were also fitted with a new research processor (SPEAR)[3] developed at the Cooperative Research Center (CRC) for Cochlear Implant and Hearing Aid Innovation in Melbourne. They used this processor for 2 to 3 weeks after which time they returned for a second visit and were retested. The subjects also participated in psychophysical studies during both visits as time permitted.

Presently all commercial sound-processing strategies for pulsatile nonsimultaneous implant systems use a bank of band-pass filters and sample the *envelope* of the signal at the output of each band. The rate at which that information is sampled is sometimes referred to as the update rate. The envelope information is then used to set stimulation levels on corresponding electrodes at fixed stimulation rates, which may or may not be the same as the update rate. This means present strategies do not preserve fine-timing information in the signals from each filter band and in fact may introduce arbitrary fine timing cues between the ears if perceptible. The SPEAR processor was in our studies programmed with a new bilateral strategy developed at the CRC. This strategy, henceforth referred to as PDT (peak derived timing), aims to preserve fine-structure ITD information by locating positive peaks in the fine timing at each filter-band output and stimulating the associated electrode for each band at times corresponding to those peaks. In addition to the new sound-processing approach, probably the most important difference between the research device and the clinical sound-processors is that the SPEAR does not use an automatic gain-control (AGC) circuit to limit high-intensity signals.

In the PDT implementation tested in these experiments, ten filter-bands with boundaries spaced approximately equally on a logarithmic scale and in total spanning approximately 250–7000 Hz were used. The boundaries for the filters were the same for each ear, but the assignment to electrodes was varied to compensate for insertion offsets between the arrays in the two ears. Because x rays were not available, the offset was estimated using a pitch-magnitude estimation task in which subjects were required to assign numbers to the pitch of electrodes activated in either ear in random order.[4] Stimulation levels were chosen to elicit comfortable loudness sensations, comparable to everyday conversational levels with the subjects' sound-processors. In order to save time, only every second electrode of the 22 electrodes available in each ear was included. Data were collected from 20 repetitions for each electrode and results were

used to determine an average offset in number of electrode-bands for comparable pitch percepts in the two ears. For subjects A–E respectively, the average offsets and standard deviations (σ) for the ten bands tested in each ear were 0.6 (σ=0.5), 1.2 (σ=1.3), 0.7 (σ=1.1), 2.5 (σ=1.5), and 0.2 (σ=1.1) bands. Since subjects A, C, and E showed offsets smaller than one electrode band, no insertion offset compensation was used in the PDT maps for them. This decision was also influenced by the consideration that, given the relatively short duration of the study, it might be advantageous to these subjects to preserve the same relative frequency mapping in the two ears as the clinical processors they had been using for some time. For subjects B and D, who showed larger pitch differences, offsets of 1 and 2 electrodes were used respectively. In those cases, for the ear with the shallower insertion, the ten filter-bands were assigned to ten electrodes using every second electrode on the array starting with the most apical one for the lowest frequency range. For the contralateral ear, the same filter-bands were used but an adjustment was made in the electrode assignment corresponding to the estimated average insertion offset.

Stimulation rates with the PDT strategy were limited to less than about 1400 Hz on any electrode to stay within the maximal overall stimulation rate of 14.4 kHz available from the implant hardware. If peaks occurred at the same time in more than one filter-band, an arbitration scheme used both amplitude and frequency information to decide which corresponding electrical pulse was to be advanced or delayed slightly to accommodate the nonsimultaneous stimulation constraint imposed by the implant system. In developing the PDT strategy, both mathematical models and real time analysis of the actual implementation on the SPEAR were used to verify that it does indeed preserve interaural fine-timing fairly well, particularly for the lower filter bands that span the important range up to about 1500 Hz. When arbitration is required to resolve temporal clashes, pulses are predominantly shifted in higher frequency channels where fine timing ITDs are not well perceived for normal-hearing listeners. For the lowest four channels (up to 1250 Hz) our simulations showed that averaged across vowel, babble, random and speech-weighted noise signals, pulses were shifted only about 3% of the time and then only by a single sample interval of about 70 μs. In comparison, clinical strategies that only extract envelopes and present that information independently at each ear at a fixed stimulation rate can in effect introduce much larger ''fine-timing'' ITD errors of up to half the stimulation interval (typically of the order of 0.5 ms). Note that a simple modification of such systems, locking the time of stimulation at the two ears together, does not provide the correct fine-timing ITD coding for any signals except those lying on the plane midway between the two ears.

Monaural thresholds (T-level) and maximal comfortable levels (C-level) for the PDT strategy were measured using constant-rate electrical pulse trains representative of the expected average stimulation rates in each filter-band. These were determined from measurements using the strategy with speech, broadband noise, and harmonic tone complexes. For each bilateral offset-adjusted pair, loudness at both T- and C-levels was compared by continuously playing left and right single-electrode stimuli in an alternating sequence. Adjustments were made to the T- and C-levels if discrepancies were found between the ears. Levels for broadband stimulation were subsequently checked with the processor in ''live mode'' by presenting noise to left and right ears in an alternating sequence and by questioning subjects about the ''centeredness'' of their own voice.

A binaural map compensating for binaural loudness summation was derived from the monaural maps using an iterative approach, starting with a reduction of all electrical T- and C-levels by about 10% of the dynamic range (in clinical level units). Adjusting both T- and C-levels by the same amount rather than either alone was done to preserve the gradient of the acoustic-to-electrical level-mapping function, which for speech in noise is likely to better preserve the same S/N for monaural and binaural listening (van Hoesel *et al.*, 2002). When assessing monaural and binaural loudness percepts, care was taken to (1) ensure left/right balance for binaural listening and (2) to instruct listeners to attend to the overall loudness in the binaural case rather than to ''hear out'' a left or right-ear contribution. The processor was programmed so that left, right, or binaural listening could be selected using a switch. Subjects listened to both the experimenter, directly in front of them, and their own voices while switching between the three listening conditions to make loudness comparisons. If necessary, further adjustments were made by shifting both T- and C-levels in the maps. All subjects reported a matched monaural and binaural loudness within one or two iterations.

For the psychophysical lateralization studies, the insertion offset estimated from the pitch task was used as an approximate starting point to select matched-place electrode pairs. Place matching was then further assessed using paired comparison of pitch to find individual bilateral electrode pairs that were well matched. Only pairs that sounded ''the same'' or were very difficult to discriminate were used in the lateralization studies. Note that when activating electrodes in the two ears, some current-spread will result and may partly counteract the need for perfect matching of place of stimulation in the two ears. Furthermore, even if exactly the same place along the cochlea on each side is stimulated, we are likely to obtain different neural response patterns in each ear due to variations in, for example, neural distributions and current flow geometries. In the lateralization experiments, electrical stimuli were generated using direct computer-control of the implants. With the system used, ITDs could be controlled with better than 3-μs accuracy. Stimuli always comprised monopolar, biphasic current-pulses with positive and negative phases that were both 25 μs in duration and separated by an ''interphase gap'' of 20 μs. Stimulation rates were varied between 50 and 800 pps. The electrode numbering convention used in this manuscript is in accordance with the clinical convention for the device, which numbers the most basal electrode band as ''1'' and the most apical band as ''22.'' To specify left and right components of each bilateral pair electrode numbers have been prefixed by ''L'' and ''R,'' respectively.

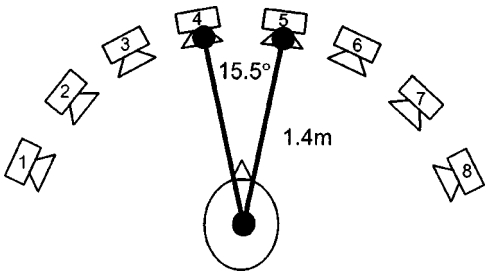


FIG. 1. Schematic diagram showing the loudspeaker array used for the localization tests.

III. LOCALIZATION

A. Methods

Localization abilities for the five subjects were tested using a loudspeaker array shown schematically in Fig. 1. The array consisted of eight loudspeakers spanning approximately the frontal 108-degree horizontal arc at 15.5-degree intervals, and at a distance of 1.4 m from the center of the head. The loudspeakers were numbered 1–8 from left to right and were set up in an anechoic room. The paradigm used was adapted from Byrne *et al.* (1998). Stimuli consisted of a sequence of four 170-ms pink-noise bursts, including 10-ms onset and offset times, and separated by 50-ms intervals. The total stimulus duration was 830 ms. Although the subjects' heads were not restrained, listeners were instructed to try to keep their heads directed towards a marker placed halfway between the two center loudspeakers. Each stimulus was presented nominally at 65 dB SPL (measured at the position of the listener's head in absentia) with ±4 dB random intensity variation in the overall level for each presentation. The intensity variation was used to eliminate any residual overall loudness variations between loudspeakers (after calibration) and reduce the monaural level cue resulting from the head-shadow. Although the range of 8 dB is comparable to the level differences observed for signals arriving from ±90° with the directional microphones used (van Hoesel *et al.*, 2002, Fig. 2), on average there is still a monaural intensity cue available. However, it was not our intention to totally eliminate monaural cues (including spectral cues). Instead, we were more interested in comparing monaural and binaural performance with signal cues that might typically be available in everyday listening. The nominal level of 65 dB SPL was chosen to avoid activation of the automatic gain control (AGC) circuits for the subjects' clinical devices. As was shown in van Hoesel *et al.* (2002), higher levels can adversely affect localization performance because of asymmetric activation of the AGCs, which distorts ILD cues. This is not an issue with the research processor because it does not include AGC circuits.

The subjects were able to see all eight loudspeakers and were required to indicate, after each stimulus presentation, which loudspeaker had sounded. No repeats were allowed and no feedback was given at any time. Each of the eight source-locations was included once in random order within a single block. Eleven blocks were combined in each test-period and were presented without any pauses delineating blocks. Data were collected from two such periods, with a 5-min break in between. Before each period, the subject was presented with two sweeps across the eight loudspeakers, during which the noise was presented in turn from loudspeaker 1 through to 8. Although subjects were unaware of the fact, the first block in each test-period was used to familiarize the listener with the task so that in total 20 repeat-presentations from each loudspeaker were included in the analysis. Whenever subjects switched between left, right, or binaural listening, they were engaged in casual conversation for about 5 min to become somewhat acclimatized to the new listening condition. For the research device, when switching between listening conditions, the loudness mapping functions were adjusted by software in the processor to compensate for binaural loudness summation. For the clinical devices, no adjustments for binaural loudness summation were made when switching between listening conditions. Left, right, and binaural listening were tested in random order across subjects using an incomplete Latin-square design.

B. Results

Figure 2 shows localization data for all five subjects. The five graphs in each column are the results for individual subjects. The first and second columns show the means and standard errors for responses to stimuli presented from each of the eight loudspeakers, with clinical and research processors, respectively. Data for both unilateral and bilateral device use are shown. The third column shows rms-errors for responses to stimuli from each individual loudspeaker, for binaural listening using clinical (circles) and research (triangles) processors. The total rms-average errors for the entire loudspeaker array are shown in Table II. Although some differences among subjects are evident, all of them show a clear advantage for the binaural condition compared to either ear alone for both clinical and research processors. Typical binaural rms errors averaged across the eight loudspeakers were around 10° for all five subjects with either processor. No learning effects were observed in the data when comparing the performance for each of the two 11-block periods for each listening condition. Monaural error scores were generally three to four times larger although considerable variation is evident across subjects, and in some cases even between ears for the same subject (subjects A and B). The binaural rms errors for individual loudspeakers (column 3 in Fig. 2) often showed better performance for loudspeakers close to the middle of the array, particularly for the research processor. The rms-averaged errors with the research processor for just the central two loudspeakers (i.e., about ±8° from the center-line) were 2.5°, 4.2°, 4.9°, 3.5°, and 11° for subjects A–E, respectively. The best performance, averaged across the entire array, was found for subject D using the research processor. This subject showed an overall rms error of only 7° when all eight loudspeakers were included, and when loudspeaker 8 was excluded, the error reduced to just 5°.[5] Because this subject was quite quick in completing the task, an additional test condition was included with the loudspeaker-array directly behind her. The data for this test are shown for subject D in column 2 of Fig. 2 by the additional gray symbols. The overall rms error in that case was around 17°.

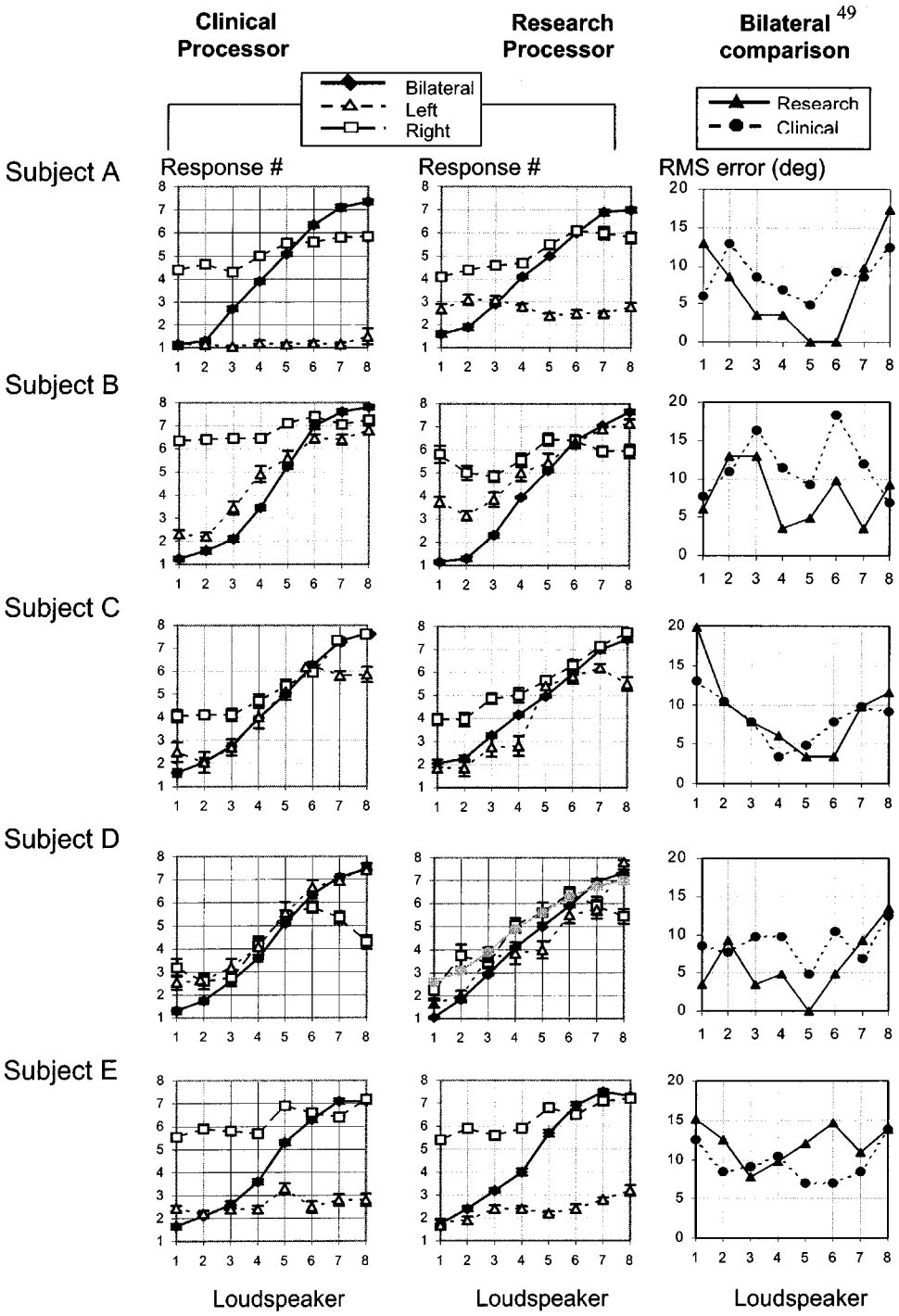


FIG. 2. Localization data for all five subjects. Columns 1 and 2 show, for clinical and research-processors, respectively, the mean results and standard errors for 20 presentations from each loudspeaker. Left, bilateral, and right ear data are shown by triangles, diamonds, and squares, respectively. In column 3, rms errors (in degrees) for bilateral device use are shown for each of the eight loudspeakers. The gray symbols and lines for subject D indicate additional test data when loudspeakers were behind the subject.

C. Discussion

For all five subjects, binaural localization abilities were clearly better than monaural abilities with the ''better ear'' (in terms of localization performance). Binaural rms errors averaged across all eight loudspeakers were about three times smaller than the better ear alone. For three of the five subjects, monaural performance in at least one ear was near chance, which is about 50° for this test configuration. The large improvement for bilateral device-use is evident with both clinical and research processors. Differences in performance between the two are minor in comparison with the differences between unilateral and bilateral device use, which may imply that the added interaural timing information with the research device did not help in this task. In any case, the comparison between devices is confounded by processor hardware variations (such as dynamic range, inherent noise considerations, and AGC circuitry), differing number of electrodes, and device listening experience. The fact that monaural performance was always much worse than binaural performance implies that residual overall intensity cues, after the 8-dB random variation in presentation levels, did not provide a robust localization cue in comparison to binaural cues. The same is true of the monaural spectral cues in the signals. It is interesting to note the increased variance in the unilateral data compared to the bilateral results for either clinical or research processors. It may be tempting to assume

TABLE II. Rost-mean square localization errors (in degrees of arc) averaged across all eight loudspeakers for left, bilateral, and right ears and for both clinical and research processors.

Subject	Left (Clinical)	Left (Research)	Bilateral (Clinical)	Bilateral (Research)	Right (Clinical)	Right (Research)
A	61	45	9	9	30	28
B	20	25	12	8	46	40
C	28	27	9	10	26	27
D	23	25	9	7	32	32
E	47	41	10	12	40	38

that this is largely because of subjects' varied abilities to associate monaural spectral and overall level cues with sound direction. However, Table II shows that at least two of these subjects (A and B) showed considerable differences in average rms error-scores between left and right ear performance. This implicates the salience of the cues rather than the ability of the subjects to associate the cues with different locations. Further studies, comparing, for example, electrode discrimination with unilateral localization abilities, are required to explore the issue.

As with normal hearing, the gradient of the ILD and ITD cues in the acoustic signal, as a function of angle of incidence, is greatest near the midline for the directional microphones used. The differences between the cues for adjacent loudspeakers are therefore larger near the middle of the array than near the end. Examining performance for just the two middle loudspeakers in the array for the pink-noise bursts, four of the five subjects showed bilateral rms-errors between 2° and 5°. This is an excellent result and implies a minimal audible angle (MAA) approaching that for normal-hearing listeners, which is close to 1° for broadband stimuli (e.g., Perrott and Saberi, 1990). It does not mean, however, that these subjects necessarily had access to the same cues as normal-hearing listeners. In the normal-hearing case the low-frequency ITD cue, when available, causes the rms-averaged array-error to vary fairly linearly with span. However, from our experiments, it seems these implant-users predominantly relied on ILD cues, which are non-monotonic beyond about 60° (Fig. 2, van Hoesel *et al.*, 2002). This means we cannot simply scale our total array rms-errors to obtain estimated errors for alternative spans.

IV. SPEECH PERCEPTION IN NOISE

A. Methods

To assess some of the speech intelligibility benefits available to these bilateral-implant users an adaptive speech reception threshold (SRT) paradigm was used. The test used BKB sentences (Bench and Bamford, 1979) presented in broadband noise that had the same long-term spectral shape as the sentences. Sentences were presented from a loudspeaker directly in front of the listeners at a fixed level of 65 dB SPL (A-weighted) in a clinical sound-booth. Although the booth was carpeted and doors and walls were paneled with ''nonreflecting'' tiles, it contained some hard reflective surfaces such as windows. The noise was presented either from the same loudspeaker as the sentences (S0/N0 condition) or else from a loudspeaker placed 90° to the left (S0/NL) or right (S0/NR) of the listener. After each sentence, listeners were asked to repeat what they had understood. Each sentence contained three or occasionally four key words. Subjects were allowed one error in the reported key words before the sentence was scored as incorrect. After each score, the noise level was adjusted.[6] If correct, the S/N was decreased, whereas after each incorrect score it was increased. The step-size used was 4 dB for the first two turning points and 2 dB after that. The last eight turning points out of ten were averaged to obtain the SRT. Because only one or two sentences in each list contained four key words rather than three, the equilibrium S/N was only marginally affected by the shift in target-criterion with four key words so that the algorithm was tracking a S/N corresponding closely to a 50% correct criterion. The process was repeated at least twice for each combination of ''ear'' (left/both/right) and noise location (NL/N0/NR). Left, right, and binaural conditions were tested according to a randomized Latin square design across subjects. For each subject with each ''ear,'' noise locations were similarly randomized. For each subject repeat measurements were made by completing the entire test sequence of nine SRT measurements in reversed order. To comply with time restrictions, SRTs were only measured with the research processors. As discussed in the Introduction and in Sec. I, binaural loudness summation was allowed for in the processor by including both unilateral and bilateral loudness mapping functions. As in the localization task, when switching between left, bilateral, and right listening conditions subjects were engaged in casual conversation for 5 min to allow brief acclimatization to the new listening condition.

B. Results

SRT estimates for the BKB sentences in noise with the research processor are shown for four subjects in Fig. 3(a). Results from subject A, who showed the worst performance (SRTs typically between 5 and 15 dB), were excluded because stable measurements could not be achieved. In each plot three groups of bar graphs show, from left to right, the SRT when noise was on the left (NL), in front (N0), and on the right (NR), respectively. The error bars indicate the standard error for the two measurements for each condition and are equal to zero if absent. For each noise location, results are shown from left to right, for listening with the left ear only (L), both ears (B), and the right ear only (R).

When both speech and noise were directly in front of the subjects (S0/N0) there is no evidence from these data that listening with both ears was any better than listening with the better ear alone. The average difference for the four subjects

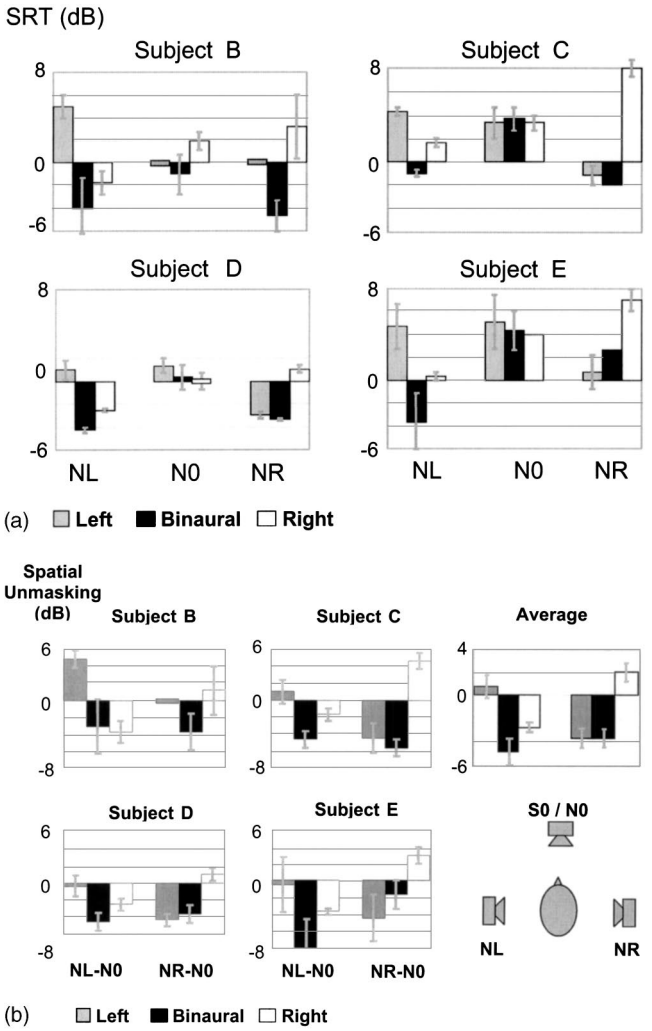


FIG. 3. (a) Mean SRTs and standard error bars for four subjects with BKB sentences in spectrally matching broadband noise. For each subject results are shown from left to right for noise arriving from the left (NL), front (N0), and right (NR), respectively. For each noise position, bar graphs are for listening with the left ear alone (L), both ears (B), and right ear alone (R), respectively. (b) Spatial unmasking data calculated by subtracting SRTs in (a) for the N0 condition from NL and NR conditions. Individual data and average for the four subjects included in the analysis (see text).

comparing bilateral device use with the better ear was 0 dB. When noise arrived from the left or right side, monaural performance ipsilateral to the noise often deteriorated, contralateral performance nearly always improved substantially, and binaural performance was usually as good or better as performance with the contralateral ear. The data for those conditions have also been replotted in Fig. 3(b), which shows the differences in SRTs when noise was either on the left (NL-N0) or right (NR-N0) side compared to when it was in front of the listeners. This indicates the amount of spatial unmasking for NL and NR conditions compared to the N0 case. A negative difference indicates improved performance and shows that when the noise was moved to the side the subject was able to tolerate more noise (with the ear contralateral to the noise or binaurally) compared to when it was directly in front. Because the data for each subject (and each ear) are normalized with respect to his or her corresponding N0 condition data, they can be averaged across subjects as shown in the rightmost set of graphs. The averaged results show that when the noise was moved to the left side, the left ear SRT worsened by about 1 dB, and right ear SRT improved by almost 3 dB. The difference between the two ears, resulting from monaural head-shadow effects, in this case is about 4 dB. Binaural performance showed an additional benefit (binaural unmasking) of about 2 dB[7] compared to the right ear. When the noise was moved to the right, the right ear SRT deteriorated by about 2 dB and the left ear improved by about 3.5 dB. The difference between ears in this case is about 5.5 dB. The bilateral condition showed no benefit over the shadowed ear in this case.

C. Discussion

The SRT data show that when the noise arrived from 90° to one side, binaural listening offered a clear head-shadow advantage irrespective of whether the noise was on the left or right. This is in agreement with the normal-hearing literature (e.g., Zurek, 1993) as well as other studies with bilateral implant users (van Hoesel and Clark, 1999; van Hoesel *et al.*, 2002; Tyler *et al.*, 2002; Gantz *et al.*, 2002 and Müller and Helms, 2002). For monaural listening, the data show that the head-shadow offers an advantage only when the noise is on the contralateral side, and can in fact be a disadvantage if ipsilateral.[8] An analysis of variance with *post-hoc* Tukey simultaneous tests was applied to the data depicted in Fig. 3(a). Results indicated that, when signal and noise were both in front of the listeners, there were no significant differences between left, right, or bilateral device-use. For noise on the left or right side, the differences in means between ipsilateral and contralateral ears (because of the monaural head-shadow) were about 5 dB and were highly significant ($P < 0.001$). The benefit we might ascribe to binaural unmasking when comparing binaural performance with the shadowed ear was less than 2 dB and was only weakly significant ($P = 0.04$). Some of the other bilateral implant studies cited in the Introduction of this paper also show varying evidence that the use of two ears can result in improvements compared to just using the ear with the better S/N but generally agree that this benefit is less than that available from the head-shadow. In our study, when speech and noise both arrived from in front of the listeners, binaural performance was comparable to that with the better ear alone. Although the data from Tyler *et al.* (2002), Gantz *et al.* (2002), and Müller and Helms (2002) indicate that some subjects showed a binaural advantage for speech in quiet or spatially coincident speech and noise, as we have discussed in the introductory section of the manuscript, varying methodologies may result in different contributions to the observed binaural benefit. In any case, we note that for normal-hearing listeners the S0/N0 SRT benefit is only about 1 or 2 dB (e.g., Bronkhorst and Plomp, 1989). Note that even in our own study where we have attempted to reduce the potential impact of listening experience by allowing a 5-min acclimatization period when switching to new listening conditions, we cannot be sure that longer exposure to monaural listening would not result in better monaural performance. If so, that would further reduce observed binaural benefits, including our estimate of binaural unmasking.

In the absence of any strong binaural unmasking effects, the main benefit to bilateral implant users for speech in noise is the ability to capitalize on the head-shadow effect when speech and noise are separated. In our study, all the subjects showed comparable post-implantation performance in the two ears, despite the selection criterion of having differences between the ears prior to implantation. However, some bilateral implantees may have one ear that performs considerably better than the other. If the performance difference for left and right ears (for S0/N0) is greater than the effect of the head-shadow, when the noise is ipsilateral to the better ear the subject in effect does not have an ear with an improved S/N. Severely asymmetric performers are therefore considerably less likely to obtain what is probably the largest speech intelligibility benefit from bilateral device use. Nevertheless there are other potential benefits to consider for those subjects, such as ensuring ''the better ear is implanted'' and the fact that localization will probably still be much improved and can act as an aid to directing attention to maximize information. The dominance of the head-shadow effect in our data also leads us to speculate about the possibility of using relatively simple algorithms with a microphone behind each ear, to automatically route signals to a single implant. However, the main problem facing such a system is determining the signal of interest (which need not necessarily be the dominant speech signal) from the total input. For a bilateral implantee, both ears can continuously be monitored and either ear can be selected at will.

V. LATERALIZATION

A. Methods

Two lateralization experiments were conducted using direct computer control of ILDs and ITDs for electrical pulse trains in each ear. The first of these was a ''lateralization-discrimination'' task requiring left/right judgments of a probe position compared to a centered reference signal and the second was a ''lateralization-position'' task in which subjects indicated absolute lateral positions on a line. Most subjects were tested with two or three matched-place, apical electrode pairs. The initial estimates for the place-matched pairs were derived from the pitch-magnitude task as described in Sec. II. A number of usually fairly apical electrode pairs were selected and a verbal report of how similar the pitch was for monaural stimulation in the two ears was requested for each pair. After obtaining similar reports for neighboring bands on one side to determine if they might result in better pitch matches, the most similar pairs were used in a pitch-matching task in which left and right ears were stimulated monaurally in a continuously alternating sequence. Between five and ten comparisons were made and after each sequence the subject was asked which ear, if any, had the higher pitch. Pairs that resulted in pitch percepts in one ear that were higher than the other for 75% or more of the comparisons were considered to be substantially different and in that case alternative pairs were explored to try to find a better match. Additional matched pairs were tested as time permitted. When introducing ILD differences, overall level was not roved. The main reason for this was the consideration that an equal increase in stimulation current in each ear, or even equal increase in terms of percentage of the dynamic range for electrodes in each ear, might result in asymmetrical contributions to loudness and/or shift lateralization percepts because of stimulating different neural distributions in the two ears. It may be possible to test for such lateral shifts for each subject before roving levels in an experiment like this but that was not feasible within our time constraints. Although theoretically the ILD cue was therefore confounded with monaural level cues in these experiments, several aspects of the tasks were designed to encourage using information from both ears. In the lateralization-discrimination task, for the smallest available ILD (0.17 dB) the level was by necessity adjusted only in one ear by a single clinical unit and was equally likely to occur on either side for each presentation. Furthermore, for subjects who could hear small ILD cues, performance for that condition was also to be compared with the case where an equal and opposite adjustment was made in the contralateral ear (ILD=0.34 dB). In the lateralization-position task, stimuli could contain either ITDs or ILDs (in random order) on each trial so that again monitoring levels in one ear only would be a poor strategy to adopt. The instructions in both tasks also explicitly stressed positional judgments, and when questioned all subjects reported that was, indeed, the cue that they were aware of monitoring. This was true whether ITDs or ILDs were being manipulated.

In the first experiment, the lateralization-discrimination task, subjects were asked to indicate whether stimuli appeared to originate from lateral positions to the left or the right of a preceding reference stimulus. The reference and probe stimuli were always on the same bilateral electrode pairs and differed either in stimulation levels at one or both ears (ILD cues) or time of stimulation (ITD cues) but not both. The reference stimulus always had zero ITD and stimulation levels were adjusted to give the most centered image[9] using an iterative approach. To obtain a centered image the electrode in one ear was stimulated at a fixed level near 85% of the dynamic range and the contralateral ear was stimulated at a number of levels, starting at 85% of its dynamic range, until the image was centered according to subjective reports. Because the same absolute current levels in each ear can produce very different loudness percepts with electrical stimulation, we refer to the centered image condition with 0-ITD as the ''0-ILD'' condition, even though actual stimulation levels may differ in the two ears. For each subject, psychometric curves were constructed for the range of ITD or ILD cues tested with each stimulus configuration tested. The just-noticeable difference (jnd) was defined as the value of the physical cue required for 75% correct responses according to the psychometric curve.[10] In order to encourage the listeners to use the lateral position of both reference and probe signals, instructions to the listeners emphasized that they should indicate relative lateral positions of target and reference sounds rather than the absolute position of the target. Nevertheless, we cannot rule out the possibility that some subjects may have ignored the reference causing the (equivalent 2AFC) jnd to be underestimated by a factor of $\sqrt{2}$. Eighty repeat presentations, half with lateralization cues to the left and half to the right, were included. Percent cor-

rect responses for stimuli with left and right cues were averaged to obtain the scores used to generate psychometric functions. The range of ILD and ITD cues studied was varied to span the range from near perfect performance, where possible, to less than 70% and was typically limited to four or five cue sizes to comply with time restraints. Each 75% threshold was estimated from a maximum-likelihood cumulative Gaussian fit to the left-right averaged data. ITD sensitivity was measured for stimulation rates over the range 50–800 pps. ILD sensitivity was measured for constant amplitude, fixed rate pulse-trains at 200 pps. Even-sized ILDs (in terms of clinical levels) were applied symmetrically by increasing the level in one ear and decreasing it by the same number of levels in the other (e.g., an ILD of four levels was generated by increasing the level in one ear by two steps and decreasing the contralateral side by two steps). Odd-sized ILD cues had a decrement in one ear that was one step larger in magnitude than the contralateral increment, except for the smallest ILD of one step, where only an increment on either side was used. All stimuli were 300 ms in duration.

In the second experiment, the lateralization-position task, subjects were required to indicate lateral positions on a line representative of the distance between the ears, by using a cursor on a computer screen. Stimuli were 50 pps fixed amplitude pulse trains that were again 300 ms in duration and contained either ILD or ITD cues. Subjects were allowed to respond with multiple positions if more than a single percept was heard. ILD variations were created by starting with the same centered reference stimulus as used in the fist experiment and stepwise increasing current levels in one ear by one clinical step (0.17 dB) and decreasing the contralateral level by two steps (0.34 dB). The asymmetrical adjustment procedure was used to allow for the wider range of ILDs explored compared to the lateralization-discrimination task and the observation that, particularly as stimulation levels approached maximal comfort levels, loudness increased more for equivalent increments (in dB). Five stimulus variations with increased levels on the left side and five with increases on the right side were included. This means that stimuli with the largest ILD cues used stimulation currents at $+0.85$ dB in one ear and -1.7 dB in the other, giving an electrical ILD of about 2.5 dB. For reference, we note that in our experience, the monaural electrical dynamic range for stimuli such as these is usually less than 10 dB. Although an ILD of one-quarter of the dynamic range or more may seem like a very large cue in normal-hearing terms, it is worth remembering that in sound-processors used in everyday life by these subjects, an acoustic input dynamic range of only 30–40 dB is typically mapped to that electrical dynamic range. Ten stimuli with ITD cues, half with left and half with right ear stimulation leading, were generated by delaying the entire stimulus pulse-train in one of the ears over the range 0–800 μs. This range was chosen to encompass that available from natural head-width delays (e.g., Blauert, 1997). All the ILD and ITD variations were included in random order a single block, which was presented 12 times in a single period without pauses delineating blocks, requiring in total 240 estimates of lateral position.

TABLE III. Psychophysical sensitivity data estimated from psychometric functions derived from lateralization responses to electrical stimuli with either ITD or ILD cues. ''Best'' and ''Worst'' ITD sensitivity for the simple 50-pps pulse-trains indicates the range of values found for a limited number of electrode place pairs tested in the time available with each subject. Subjects A, B, and C also repeated the lateralization test for ITDs to the entire signal for 800-pps pulse-trains modulated at 50 Hz using the same electrodes that gave the ''Best'' performance for the unmodulated 50-pps pulse-trains. Subject A also repeated the test for delays in just the 50-Hz modulation, preserving 800-pps coincident ''carriers'' in both ears. ILDs (dB) are described in terms of stimulation current level differences with respect to the ''centered image'' condition.

	ITD 50pps (μs)		ITD 800mod 50 (6.8 dB moddepth)		
Subject	Best	Worst	Delay all	Delay env only	ILD (dB)
A	90	100	120	290	<0.17
B	120	170	160		0.34
C	90	245	130		<0.17
D	150	180			0.51
E	180	250			0.68

B. Results

Table III shows the sensitivity to ITD and ILD cues estimated from the lateralization-discrimination experiment. ITD sensitivity at 50 pps is shown on the left side (second column) and ILD sensitivity for the 200-pps stimuli is shown on the far right-hand side of the table. For the ITD measurements at 50 pps, results are shown for the best- and worst-case jnd values found for the electrode pairs examined in the available time. The electrode pairs displaying the best ITD sensitivity at 50 pps were also used to measure ILD sensitivity and to examine the effect of higher pulse-rates on ITD sensitivity as shown in Fig. 4. The higher rates used were 200 and 800 pps for subjects A-D, and 100 and 200 pps for subject E.[11] Note that at 800 pps the interval between pulses is 1.25 ms and delays beyond about 600 μs convey ambiguous ongoing delay information. For that reason ITDs used were limited to the range 0–400 μs. Although the number of

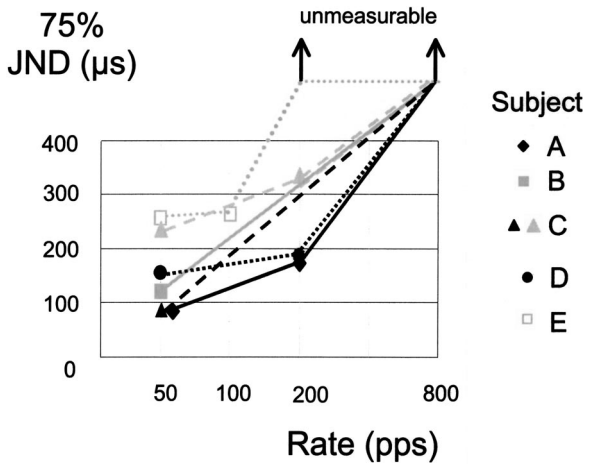


FIG. 4. Sensitivity to ITDs (75%, 2S2I; see text) as a function of rate of stimulation for unmodulated pulse-trains. The upward arrows indicate that at the rate indicated subjects were unable to obtain 75% correct for 400-μs ITDs and that the jnd is expected to be larger than 400 μs. Subject C repeated the test for two different bilateral electrode-pairs, as indicated by the black and gray triangles.

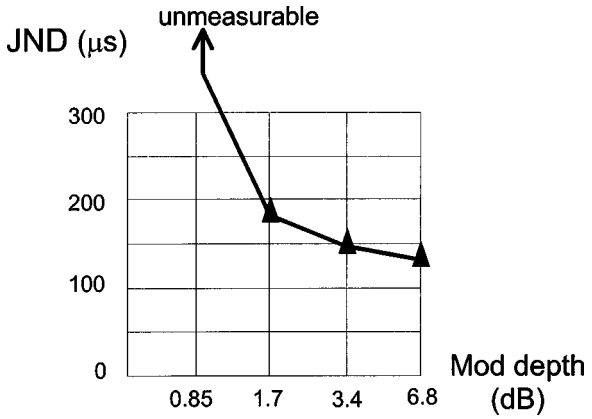


FIG. 5. Additional data from subject A. Sensitivity to ITDs applied to the entire stimulation waveform as a function of modulation depth of the stimulation current (dB) for an 800-pps pulse-train, log-sinusoidally modulated at 50 Hz. At the fairly shallow modulation of 0.85 dB, the subject was unable to obtain 75% correct for ITDs of 400 μs and the upward arrow indicates that the jnd was probably above 400 μs.

rates that could be tested was limited, the results in Fig. 4 clearly demonstrate substantially decreased sensitivity to ITDs with the higher pulse-rates for all subjects. The upward pointing arrows for some data points are intended to convey that at those rates, subjects were unable to consistently lateralize percepts on the basis of ITDs up to 400 μs and that the psychometric functions did not rise above the 75% correct criterion we have used to determine our jnd's. Presumably, the jnd in those cases lies somewhere above 400 μs.

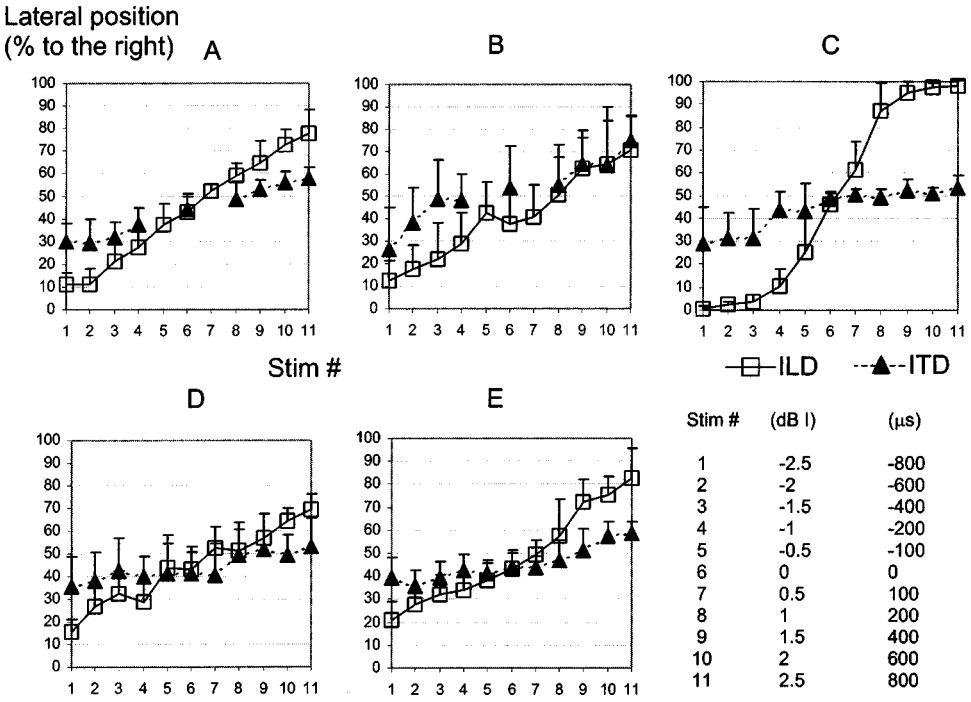
For subjects A, B, and D additional time was available to measure sensitivity to ITDs for the entire signal with pulse-trains at 800 pps modulated at 50 Hz. These results are also shown in Table III. Modulation was applied using sinusoidal attenuation of the levels (in clinical units) and resulted in a peak attenuation of 6.8 dB in stimulation current. The table includes an additional column for subject A, who was able to conduct the task quickly, and repeated the test with ITDs applied to just the modulating ''envelope'' of signal with carrier pulses held at 0 ITD.

Figure 5 shows further ITD jnd's with subject A for delays to the entire modulated stimulus (i.e., both envelope and fine-timing delayed equally) when the modulation depth was varied. Peak stimulation levels were kept identical for the different depths and no attempt was made to compensate for decreases in overall loudness with the increased modulation. It is clear that as the modulation depth of the 50-Hz modulator increased, the ITD sensitivity for the 800-pps pulse-train improved, and when it was reduced to about five clinical steps (0.85 dB) the subject could no longer lateralize consistently on the basis of 400-μs ITDs, as was the case for the unmodulated signal at 800 pps.

Figure 6 shows the results from the lateralization-position experiment in which subjects indicated lateral positions for stimuli containing ITDs or ILDs. For ITDs up to 800 μs and ILD cues of up to 2.5 dB, the data generally show a more substantial shift of the lateral position resulting form ILDs than ITDs. The percepts produced by ILDs spanned 55% to 100% of the available range from extreme left to right whereas the ITDs only shifted images over a range of 15%–45%.

C. Discussion

The lateralization data from both experiments generally show consistently good sensitivity to level cues at the two ears but variable sensitivity to ITD cues. In the lateralization-discrimination task two of the subjects, A and C, in fact showed near perfect identification of level changes as small as 0.17 dB, the smallest available adjustment. This means we could not usefully compare their performance for unilateral 0.17-dB adjustments with bilateral 0.17 dB equal and opposite adjustments. Two other subjects (B and D), however,



Stim #	(dB I)	(μs)
1	-2.5	-800
2	-2	-600
3	-1.5	-400
4	-1	-200
5	-0.5	-100
6	0	0
7	0.5	100
8	1	200
9	1.5	400
10	2	600
11	2.5	800

FIG. 6. Laterlaization-position data (mean and SD) for randomized presentation of 300-ms stimuli with either ILD or ITD cues. Eleven stimuli contained ILDs and 11 contained ITDs as indicated in the table at the bottom right. Each stimulus configuration was presented 20 times. Most subjects show a considerably larger span of lateral positions for ILDs than ITDs.

both showed improvements in comparing the bilateral adjustments with the unilateral adjustment. Their percent-correct scores for 80 presentations with unilateral adjustments were 62% and 50% whereas for the bilateral adjustment, these improved to 76% and 60%, respectively, which shows these subjects must have been using information from both ears. Subject E showed substantially worse ILD sensitivity and was not tested with these small ILD cues.

The data also show that ITDs produced discernable lateralization shifts only when stimuli contained low-frequency information below a few hundred Hz, either by using a low pulse-rate, or otherwise by modulating higher pulse-rates with a low-frequency modulator. For electrical signals containing low-rate information, most measurements of sensitivity to ITDs were around 100–150 μs. This is better than those found previously in Melbourne (van Hoesel and Clark, 1997; van Hoesel *et al.*, 2002) and better than those reported for a subject studied by Lawson *et al.* (1998)[1] but still appreciably larger than the 20 μs or so that can be achieved with normal-hearing listeners for narrow-band noise centered below 1500 Hz (Gabriel *et al.*, 1992). More recent reports from Lawson *et al.*[12] show lateralization effects resulting form ITDs as small as 50 μs but detailed descriptions of these experiments allowing comparisons with our jnd data have not yet been published. The ITD sensitivity found in our experiments so far is in fact probably closer to that found for high-frequency amplitude modulated waveforms with normal-hearing listeners, which is often in the range of one to several hundred microseconds (e.g., Henning, 1974). When low-rate information was absent in the electrical stimuli, ITD sensitivity was very poor. For 800-pps unmodulated pulse-trains, none of our subjects could correctly identify lateral shifts caused by ITDs of up to 400 μs. This is in contrast to normal-hearing subjects who with sinusoids in fact show best sensitivity to ITDs at about 800–1000 Hz (Klumpp and Eady, 1956).

For those subjects who were tested with modulated waveforms, sensitivity with 800-pps carriers was substantially improved when a deep 50-Hz modulation was applied to the stimulus and the entire waveform was delayed. For subject A it was found that delaying the modulation envelope only resulted in poorer ITD sensitivity than delaying the entire waveform. This may appear to be conflicting with the finding from the unmodulated case that shows at 800 pps ongoing fine-timing delays do not contribute substantially to lateralization shifts. However, if we consider that delaying just the envelope and not the carrier means having to ''resample'' the envelope we see that, particularly when the modulator period is an integer multiple of the carrier period, this can, in fact, introduce an unwanted ILD which may disrupt the lateralization mediated by the ITD. This explanation is supported by preliminary unpublished data from the first author's laboratory.

One of the best sensitivities measured was around 90 μs for subject C with the most apical electrodes on the arrays. Measurements with the same subject, but for electrodes that were located about five bands in the basal direction, showed ITD sensitivity of almost 200 μs. However, for other subjects, the best performance found was not necessarily for more apical electrode pairs. It is possible that monaural pitch matching to select place-matched electrodes did not result in optimal pairings of electrodes for binaural ITD sensitivity. However, as mentioned in Sec. II, it seems likely that small variations of one or two bands will have little impact on binaural sensitivity because of current-spread with electrical stimulation. The decreased ITD sensitivity as rates increased was observed for all five subjects. The mechanism governing this behavior is not yet clear to us, but may well be related to if not the same as the one responsible for monaural rate-pitch saturation after several hundred Hz often observed with cochlear implant users (e.g., Tong and Clark, 1985).

The lateralization-position data in Fig. 6 show that for all subjects lateralization shifts resulting from level cues over the range ± 2.5 dB were larger than those resulting from ITD cues over the range ± 800 μs. Only subject B showed shifts for ITDs approaching those for the level variations. However, this subject also showed rather large variance in his responses. Although the ITD range tested probably approximates the range of cues potentially available to subjects when using sound-processors in everyday situations, the chosen ILD range of 2.5 dB may overestimate that somewhat. However, if that is so, it is interesting that only subject C shows saturated lateralization responses for a number of the stimuli with larger ILD cues.

VI. CONCLUSIONS

The results from these experiments show that bilateral cochlear implants can offer important advantages for localization (in quiet) and for speech understanding when interfering noise arrives from another direction. The localization study shows that subjects can localize considerably more accurately using two devices than one. The rms-averaged localization data for an eight-loudspeaker array spanning the frontal 108° arc showed monaural rms errors between 20° and 40° with the better-localizing ear for each of the subjects tested. Binaural errors were typically three times smaller in comparison. Binaural performance for individual loudspeaker positions varied with position, and for loudspeakers close to the middle of the array rms errors were under 5° for most of the subjects. Because of the non-monotonic ILD cue with loudspeaker position, data averaged across the entire array must be treated with caution and are probably inappropriate when comparing across experiments with differing loudspeaker configurations. The speech data collected with the research processor and the PDT strategy indicate a substantial and robust head-shadow advantage. This is in agreement with other studies to date using clinical devices and allows bilateral implantees to attend to the ear with the better S/N irrespective of whether the noise is on the left or the right. In contrast, for a unilateral device with a single microphone at the ear, the head-shadow is likely to be beneficial when the noise is contralateral to the microphone but can actually be a disadvantage when the noise is ipsilateral to it. The speech data also exhibit some evidence of binaural unmasking, but the effect is comparatively smaller and less robust. Although subjects were allowed 5 min of casual conversation to acclimatize to new listening conditions when switching between unilateral and bilateral device use, we

cannot be sure that considerably more experience with the monaural condition would not further decrease the estimated amount of binaural unmasking. The dominance of the head-shadow effect for speech in noise raises the concern that bilateral implant users who show strongly asymmetric speech performance for the two ears may not obtain large benefits in this regard. Nevertheless, assuming the localization benefit is still available, there is good reason to believe that bilateral device use can assist in everyday communication by improving the ability to direct attention to the appropriate location in more difficult environments. Given the psychophysical findings that show strong effects of ILD adjustments on lateralization but only moderate and variable effects of ITDs the lack of strong binaural unmasking effects is perhaps not surprising. Although most subjects demonstrated sensitivity to ITDs in the range 100–150 μs, this was found only when stimuli contained low-frequency information. Stimuli with pulse rates above a few hundred pulses per second, unless modulated at a low rate of sufficient depth, showed extremely poor ITD sensitivity. Further work investigating in particular binaural temporal interaction with electrical stimulation is required to determine whether temporal aspects of the electrical signals at the two ears can offer added advantages to what appear to be predominantly level-derived benefits observed so far. This is of relevance to determine whether bilateral implant users are likely to need new sound-coding strategies, such as PDT, which better preserve fine-timing ITD information to benefit optimally from bilateral implantation.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the participation of the five bilateral implant users in these studies and the valuable comments on a previous draft of this manuscript by Leslie Bernstein, Quentin Summerfield, and one additional anonymous reviewer. In Iowa, we are thankful for assistance from John Preece, Bruce Gantz, Jay Rubenstein, Ruth Henry, Mary Lowder, Shelley Witt, Marla Ross, and Mark Schmidt. In Melbourne, we particularly wish to acknowledge helpful discussions with Andrew Vandali during the development of the PDT sound-processing strategy and his contributions in the implementation on the Spear processor, as well as Mark Harrison and Leslie Whitford. This work was supported by the Australian Government through their funding of Cooperative Research Center for Cochlear Implant and Hearing Aid Innovation; in part by Research Grant No. 2 P50 CD 00242 (awarded to Dept. of Otolaryngology—Head and Neck Surgery, University of Iowa) from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health; and Grant No. RR00059 from the General Clinical Research Centers Program, Division of Research Resources, NIH.

[1]The 1998 study by Lawson *et al.* used a 75% criterion in a 2S1I study requiring identification of stimuli with left-first or right-first cues. In accordance with our discussion in Sec. V A, the jnd's quoted in the Lawson paper should be multiplied by about 1.4 for comparison with our results if we assume that the subjects in our studies were indeed attending to the reference stimulus as instructed.

[2]Note that unlike with normal hearing, the same signal presented to the two ears can sound somewhat different with bilateral implants as a result of place variations of electrodes and excitation geometry in the two ears.

[3]In conjunction with a Cochlear-HS8 microphone behind each ear.

[4]This procedure was adopted from van Hoesel and Clark (1997), and for the two bilateral implantees they tested showed good agreement with the *average* insertion offset determined from radiological data.

[5]Note in Fig. 2 for this subject with the research device, loudspeaker 8 is the only one that shows a mean response that deviated significantly from the correct-response diagonal.

[6]In our initial runs, we adjusted the signal rather than the noise level. However, this occasionally resulted in unreasonably loud presentation levels when subjects could not understand sentences (even in moderate noise levels). For that reason we decided to alter the noise level instead in our tests.

[7]Note that the binaural unmasking estimates were calculated from raw scores for NL or NR rather than spatial unmasking data. When the difference between contralateral and bilateral spatial unmasking is calculated, in addition to binaural unmasking, an additional term equal to the difference between the unilateral and bilateral performance for the N0 condition (a ''diotic'' benefit) is obtained.

[8]The unilateral spatial unmasking *disadvantage* was not significant at $p = 0.05$ (ANOVA) for NL but was significant at $p = 0.012$ for NR.

[9]Note that the most centered image did not always result from simple combination of monaural stimuli that had been loudness matched by adjusting levels for alternating presentation to the two ears. Although the difference between levels for centered percepts and monaural loudness balance was not studied in detail, casual observation showed variations up to about 0.5 dB (three clinical steps).

[10]By taking this approach, we have assumed that the listeners used the lateral position of both reference and probe percepts, and treated the experiment as a discrimination task between the two [cf. Hartmann and Rakerd (1989) for a parallel discussion of the MAA paradigm]. If so, then our 75% criterion corresponds to the same criterion in a standard two-alternative forced-choice paradigm (with two stimuli, in our case one reference and one target, in two intervals: 2S2I), and our jnd estimates reflect the ability of listeners to discriminate between two signals separated by a cue size equal to the difference between the two. If, on the other hand, the listeners ignored the reference information and treated the experiment as an identification task, then they were effectively listening to just a single presentation interval (the second interval) and were required to decide whether the stimulus contained a left or right cue (two stimulus options in a single interval, 2S1I). The measured jnd in that case reflects the ability to identify which of two stimuli—separated by *twice the cue size* (because left and right cues are equal and opposite in sign)—was presented. Note that the variance of the decision statistic is in that case reduced by a factor of $\sqrt{2}$ because we are concerned with a single internal distribution rather than a difference between two distributions. The net effect is that for the 2S1I case the 75% criterion gives a jnd estimate that is smaller than the 2S2I task by a factor of $2/\sqrt{2}=\sqrt{2}$.

[11]For this subject, who had some trouble with the task, the slightly better performing place-matched pair (of only two in total) was not tested until after the rate experiment so that in the rate experiment he was tested with the pair yielding poorer performance.

[12]Twelfth Quarterly Progress Report for contract 98-01, DC8-2105 at http://npp.ninds.nih.gov/ProgressReports

Bench, J., and Bamford, J. (eds.) (**1979**). *Speech-hearing Tests and The Spoken Language of Hearing Impaired Children* (Academic, London).

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization*, revised edition (MIT, Cambridge, MA).

Bronkhorst, A. W., and Plomp, R. (**1988**). ''The effect of head-induced interaural time and level differences on speech intelligibility in noise,'' J. Acoust. Soc. Am. **83**, 1508–1516.

Bronkhorst, A. W., and Plomp, R. (**1989**). ''Binaural speech intelligibility in noise for hearing-impaired listeners,'' J. Acoust. Soc. Am. **86**, 1374–1383.

Byrne, D., Sinclair, S., and Noble, W. (**1998**). ''Auditory localization for sensorineural hearing losses with good high-frequency hearing,'' Ear Hear. **19**, 62–71.

Carhart, R. (**1965**). ''Monaural and binaural discrimination against competing sentences,'' Int. Audiol. **4**, 5–10.

Dirks, D. D., and Wilson, R. H. (**1969**). ''The effect of spatially separated sound sources on speech intelligibility,'' J. Speech Hear. Res. **12**, 5–38.

Gabriel, K., Koehnke, J., and Colburn, H. S. (**1992**). ''Frequency depen-

dence of binaural performance with impaired binaural hearing,'' J. Acoust. Soc. Am. **91**, 336–347.
Gantz, B. J., Tyler, R. S., Rubinstein, J. T., Wolaver, A., Lowder, M., Abbas, P., Brown, C., Hughes, M., and Preece, J. P. (**2002**). ''Binaural cochlear implants placed during the same operation,'' Otol. Neurotol. **23**, 169–180.
Hartmann, W. M., and Rakerd, B. (**1989**). ''On the minimum audible angle—A decision theory approach,'' J. Acoust. Soc. Am. **85**, 2031–2041.
Hausler, R., Colburn, S., and Marr, E. (**1983**). ''Sound Localization in Subjects with Impaired Hearing,'' Acta Oto-Laryngol., Suppl. **400**, 1–62.
Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (**1999**). ''Speech Intelligibility and localization in a multi-source environment,'' J. Acoust. Soc. Am. **105**, 3436–3448.
Henning, G. B. (**1974**). ''Detectability of interaural delay in high-frequency complex waveforms,'' J. Acoust. Soc. Am. **55**, 84–90.
Hirsh, I. J. (**1950**). ''The relation between localization and intelligibility,'' J. Acoust. Soc. Am. **22**, 196–200.
Klumpp, R. G., and Eady, H. R. (**1956**). ''Some measurements of interaural time difference thresholds,'' J. Acoust. Soc. Am. **28**, 859–860.
Lawson, D. T., Wilson, B. S., Zerbi, M., van den Honert, C., Finley, C. C., Farmer, Jr., J. C., McElveen, J. T., and Rousch, P. A. (**1998**). ''Bilateral cochlear implants controlled by a single speech processor,'' Am. J. Otol. **19**, 758–761.
Licklider, J. C. R. (**1948**). ''The influence of interaural phase upon the masking of speech by white noise,'' J. Acoust. Soc. Am. **20**, 150–159.
Long, C. (**2000**). ''Bilateral Cochlear Impants: Basic Psychophysics,'' Ph.D. dissertation, MIT, Cambridge, MA.
MacKeith, N. W., and Coles, R. R. (**1971**). ''Binaural advantages in hearing of speech,'' J. Laryngol. Otol. **85**, 213–232.
Müller, J., Schon, F., and Helms, J. (**2002**). ''Speech Understanding in Quiet and Noise in Bilateral Users of the MED-EL COMBI 40/40+ Cochlear Implant System,'' Ear Hear. **23**(3), 198–206.
Peissig, J., and Kollmeier, B. (**1997**). ''Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal hearing and impaired listeners,'' J. Acoust. Soc. Am. **101**, 1660–1670.
Perrot, D. R., and Kourosch, S. (**1990**). ''Minimum audible angle thresholds for sources varying both in elevation and azimuth,'' J. Acoust. Soc. Am. **87**, 1728–1731.
Rayleigh, L. (**1907**). ''On our perception of sound direction,'' Philos. Mag. **13**, 214–232.
Searle, C. L., Braida, L. D., Davis, M. F., and Colburn, H. S. (**1976**). ''Model for auditory localization,'' J. Acoust. Soc. Am. **60**, 1164–1175.
Stevens, S. S., and Newman, E. B. (**1936**). ''Localization of actual sources of sound,'' Am. J. Psychol. **48**, 297–306.
Tong, Y. C., and Clark, G. M. (**1985**). ''Absolute identification of electric pulse rates and electrode positions by cochlear implant patients,'' J. Acoust. Soc. Am. **77**, 1881.
Tyler, R. S., Gantz, B. J., Rubinstein, J. T., Wilson, B. S., Parkinson, A. J., Wolaver, A., Preece, J. P., Witt, S., and Lowder, M. W. (**2002**). ''Three-month results with bilateral cochlear implants,'' Ear Hear. **23**(1, Suppl.), 80S–89S.
van Hoesel, R. J. M., and Clark, G. M. (**1995**). ''Fusion and lateralization study with two binaural cochlear implant patients,'' Ann. Otol. Rhinol. Laryngol. Suppl. **104**(9), 233–235.
van Hoesel, R. J. M., and Clark, G. M. (**1997**). ''Psychophysical studies with two binaural cochlear implant subjects,'' J. Acoust. Soc. Am. **102**, 495–507.
van Hoesel, R. J. M., and Clark, G. M. (**1999**). ''Speech results with a bilateral multichannel cochlear implant for spatially separated signal and noise,'' Aust. J. Audiol. **21**, 23–28.
van Hoesel, R., Ramsden, R., and O'Driscoll, M. (**2002**). ''Sound-direction identification, interaural time delay discrimination and speech intelligibility advantages in noise for a bilateral cochlear implant user,'' Ear Hear. **23**(2), 137–149.
van Hoesel, R. J. M., Tong, Y. C., Hollow, R. D., and Clark, G. M. (**1993**). ''Psychophysical and speech perception studies: a case report on a bilateral cochlear implant subject,'' J. Acoust. Soc. Am. **94**, 3178–3189.
Wightman, F., and Kistler, D. J. (**1992**). ''The dominant role of low-frequency interaural time differences in sound localization,'' J. Acoust. Soc. Am. **91**, 1648–1661.
Zurek, P. M. (**1993**). ''Binaural advantages and directional effects in speech intelligibility,'' in *Acoustical Factors Affecting Hearing Aid Performance*, 2nd ed., edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston).

Auditory spatial discrimination by barn owls in simulated echoic conditions[a)]

Matthew W. Spitzer,[b)] Avinash D. S. Bala, and Terry T. Takahashi
Institute of Neuroscience, University of Oregon, Eugene, Oregon 97403

(Received 11 December 2001; revised 20 December 2002; accepted 20 December 2002)

In humans, directional hearing in reverberant conditions is characterized by a ''precedence effect,'' whereby directional information conveyed by leading sounds dominates perceived location, and listeners are relatively insensitive to directional information conveyed by lagging sounds. Behavioral studies provide evidence of precedence phenomena in a wide range of species. The present study employs a discrimination paradigm, based on habituation and recovery of the pupillary dilation response, to provide quantitative measures of precedence phenomena in the barn owl. As in humans, the owl's ability to discriminate changes in the location of lagging sources is impaired relative to that for single sources. Spatial discrimination of lead sources is also impaired, but to a lesser extent than discrimination of lagging sources. Results of a control experiment indicate that sensitivity to monaural cues cannot account for discrimination of lag source location. Thus, impairment of discrimination ability in the two-source conditions most likely reflects a reduction in sensitivity to binaural directional information. These results demonstrate a similarity of precedence effect phenomena in barn owls and humans, and provide a basis for quantitative comparison with neuronal data from the same species.
[DOI: 10.1121/1.1548152]

PACS numbers: 43.66.Qp, 43.66.Pn, 43.66.Gf [LRB]

I. INTRODUCTION

The barn owl (*Tyto alba*) is a nocturnal predator that makes use of well-developed acoustic and neuronal specializations to localize and capture prey. In a natural environment, the task of localizing a target sound source is complicated by reflections that convey conflicting directional information. Nevertheless, barn owls, like many other animals, are able to localize sounds in echoic conditions. Numerous human psychophysical studies have addressed spatial hearing abilities using stimuli that simulate a direct sound source and its reflection. In general, the results indicate that directional information conveyed by the direct (leading) source dominates perception, and that listeners are relatively insensitive to directional information conveyed by reflections (lagging sources). Perceptual dominance of the leading sound is commonly referred to as the ''precedence effect.'' Psychophysical studies of the precedence effect have typically measured one of three basic phenomena (Litovsky *et al.*, 1999). At short delays between leading and lagging sounds, subjects report ''perceptual fusion,'' meaning that the direct sound and reflection are heard as a single acoustic event (reviewed in Blauert, 1997). ''Localization dominance'' refers to the finding that information conveyed by the direct sound dominates perception of the location of the fused image, although the actual perceived location is also influenced to a lesser extent by the location of the lag source (Wallach *et al.*, 1949; Shinn-Cunningham *et al.*, 1993). Studies addressing ''lag discrimination suppression'' have shown that a listener's ability to detect changes in directional attributes of a lagging source is impaired relative to that for single or leading sources (Zurek, 1980; Perrott *et al.*, 1989; Litovsky and Macmillan, 1994). The latter two phenomena indicate that, despite the perceptual dominance of the leading source, a representation of the lagging source is retained, and that this representation contributes to perception.

Behavioral data suggest that precedence effects occur in a wide variety of species, including barn owls (Keller and Takahashi, 1996); rats (Kelly, 1974); cats (Cranford, 1982); and crickets (Wyttenbach and Hoy, 1993). In contrast to human psychophysical studies, previous animal behavioral studies have used lateralization tasks, in which the ability to distinguish between left-leading and right-leading sound pairs is measured as a function of the delay between leading and lagging sources. In all species tested, lateralization performance is very good for intermediate delay values (owl: 2 to 10 ms; cat: 0.3 to 3 ms; rat: 0.5 to 4 ms) and decreases at shorter or longer delays. Thus, at intermediate delays the presence of the lag does not interfere with the ability to lateralize the lead source. At longer delays it is presumed that the lag source becomes audible as a separate event, causing a decline in performance as subjects begin to lateralize both lag and lead sources. The range of delays at which optimal performance occurs in all animals, except crickets, is similar to the optimal range (1 to 5 ms) for discrimination of the interaural time difference of a leading sound by humans (Litovsky and Shinn-Cunningham, 2001). The fact the precedence effect occurs across classes, and even phyla, indicates that it is a fundamental process in spatial hearing.

Physiological studies have documented neuronal corre-

[a)]Portions of this work were presented in ''Behavioral measurement of a precedence effect in barn owls,'' Society for Neuroscience Conference, New Orleans, LA, November 2000, and ''Auditory spatial resolution in the barn owl under echoic and anechoic conditions,'' Acoustical Society of America Meeting, Chicago, IL, June 2001.

[b)]Electronic mail: spitzer@uoneuro.uoregon.edu

lates of the precedence effect in a variety of animal models. Single-unit studies have addressed the processing of lead–lag sound pairs by neurons that encode spatial cues (interaural time difference, ITD) at several stages of the mammalian auditory pathway (Yin, 1994; Fitzpatrick *et al.*, 1995; Litovsky and Yin, 1998; Fitzpatrick *et al.*, 1999), and by space-tuned neurons in the barn owl's auditory midbrain (Keller and Takahashi, 1996). A common finding of these studies is that the response to a stimulus at a neuron's best spatial location or ITD is reduced when that stimulus is preceded, at an interval of a few milliseconds, by a stimulus at another location or ITD. Thus, at short delays the neuronal representation of the ITD, or spatial location, of a lagging sound is at least partially attenuated. It is unclear how such neuronal effects relate to precedence phenomena at the behavioral level, as comparable measures of neuronal and behavioral responses have yet to be obtained in the same species.

The present study was designed to provide behavioral measures of the precedence effect in barn owls that can be compared directly to similar measures derived from neuronal data. A spatial discrimination paradigm was used to facilitate such a comparison, as both behavioral and neuronal discrimination performance can be quantified using equivalent metrics derived from signal detection theory (Britten *et al.*, 1992). An added advantage of spatial discrimination paradigms is that it is possible to address processing of both lead and lag sources explicitly (Perrott *et al.*, 1989; Litovsky and Macmillan, 1994; Litovsky, 1997). In addition to providing a basis for comparison with physiological data, results of this study are relevant to understanding the generality, or lack thereof, of precedence phenomena in different species. This is, to our knowledge, the first study to measure spatial discrimination of lead and lag sources in any nonhuman species.

II. EXPERIMENT 1: SPATIAL DISCRIMINATION OF SINGLE, LEAD, AND LAG SOURCES

The first experiment measured the ability of owls to reliably detect changes in the azimuth of single, leading, or lagging sound sources. Previous human studies using free-field stimulus presentation have demonstrated that minimum audible angles (MAAs) for lag sources are substantially larger than those for single sources (Perrott *et al.*, 1989; Litovsky and Macmillan, 1994; Litovsky, 1997), while lead source MAAs are intermediate between those for single and lag sources (Litovsky and Macmillan, 1994; Litovsky, 1997). A recently developed discrimination paradigm (Bala and Takahashi, 2000), based on habituation and recovery of the pupillary dilation response (PDR), was used to provide comparable measurements from barn owls.

A. Methods

Subjects were three adult barn owls from a captive breeding colony, all of which had previously served as subjects in similar experiments. Each owl had a steel headplate surgically attached to its skull. Following recovery from surgery, owls were gradually acclimatized to being restrained in the experimental apparatus in three or four sessions over several weeks (Bala and Takahashi, 2000). During experiments, owls were restrained in a cloth jacket, and supported by a metal frame. The position of the owl's head was fixed in a constant position relative to a speaker array by attaching the headplate to a holder mounted on the supporting frame [Figs. 1, 2(A)]. Once acclimatized, owls would lay quietly in the experimental apparatus without struggling throughout test sessions with durations of 1 to 2 h. During test sessions, owls were continuously monitored using an infrared sensitive video camera.

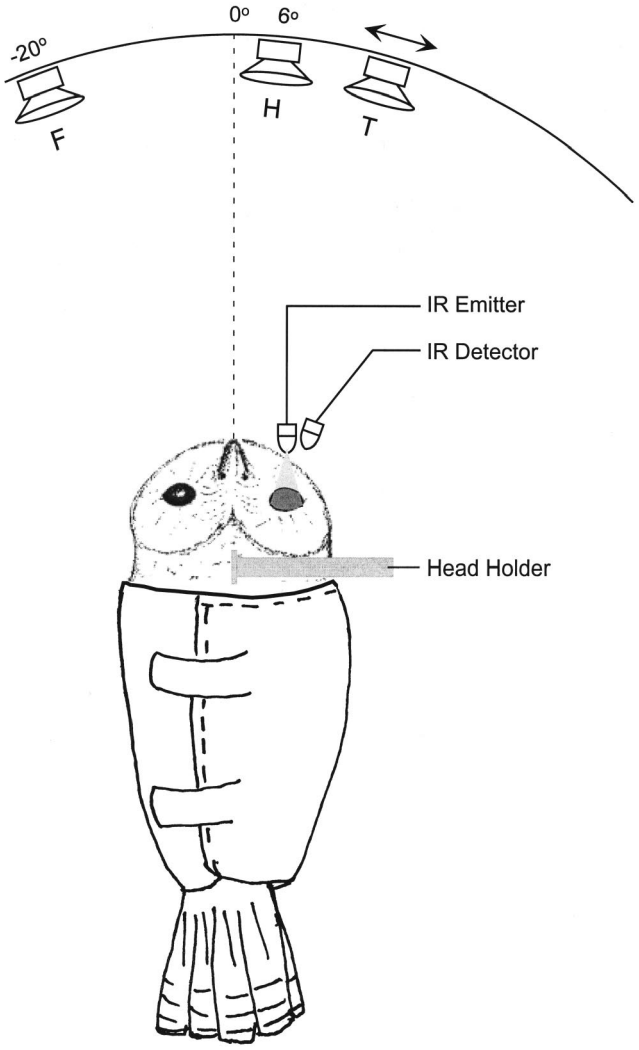


FIG. 1. Apparatus used to test spatial discrimination. Owls were restrained with their heads held fixed relative to the sound sources. Stimuli were presented from one or two of three speakers mounted on a hoop. To measure *single-source* discrimination, the owl was habituated to repeated presentation of sound from the speaker at 6° azimuth (H) and tested with sounds from a speaker to the right (T), the azimuth of which was varied between sessions. In the *lead-source* condition, the owl was habituated to a lead sound from the speaker at 6°, followed at a 3-ms delay by a lag sound from the fixed speaker at −20° (F). On test trials the location of the lead sound was changed. In the *lag-source* condition, the owl was habituated to a lead sound from the speaker at −20°, followed at 3 ms by a lag sound from the speaker at 6°. On test trials the location of the lag sound was changed. Pupillary responses were monitored using infrared emitting and detecting diodes [see Fig. 2(A)].

Experiments were conducted in a double-wall, sound-isolating anechoic chamber (Industrial Acoustics Company, IAC). The inner walls, ceiling, and floor were lined with 6-in. Metadyne 150 (IAC) anechoic absorption cones leaving

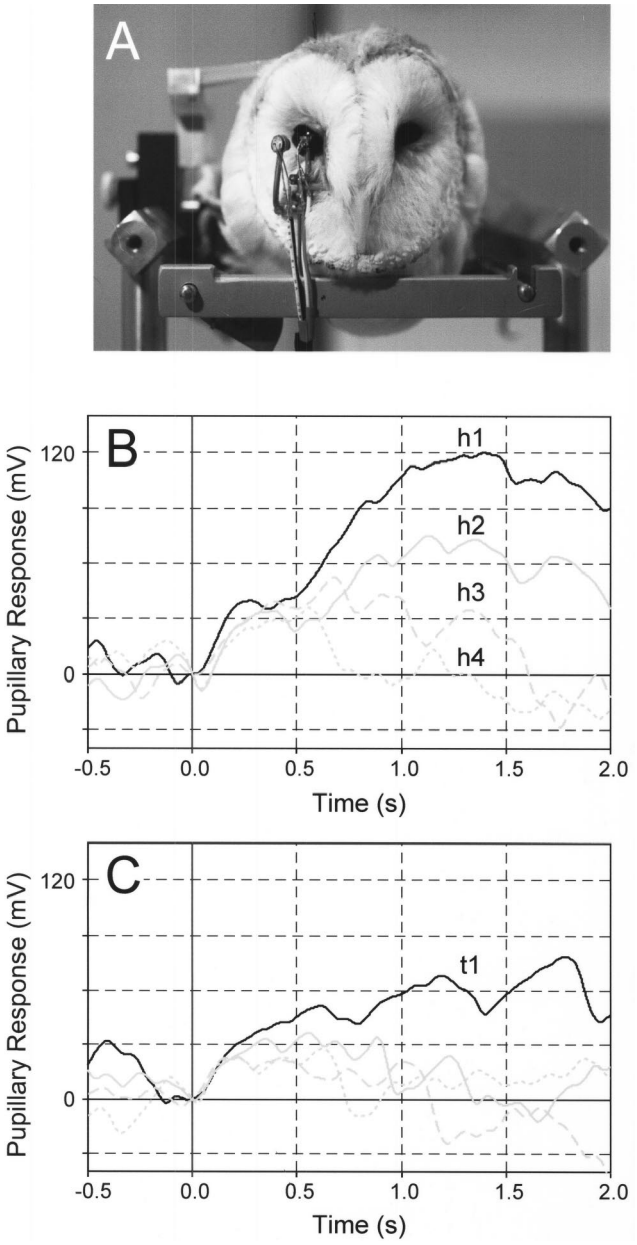


FIG. 2. Measurement of acoustically evoked pupillary dilation response (PDR). (A) Positioning of the pupillometer. (B) Initial habituation of the PDR in owl 882. Pupillary responses, measured as the voltage output of the pupillometer, have progressively diminishing amplitudes on the first four habituating trials of a test session (h1 through h4). The habituating stimulus is a lead–lag pair of 25-ms noise bursts (delay=3 ms), with the lead source located at 6° azimuth. Stimulus onset occurs at time=0 s. (C) Recovery from habituation. On the first test trial (t1, overall trial number 160), the location of the lead sound is changed by 20°. The amplitude of the evoked PDR (black line) is greater than those evoked on the three preceding habituating trials (gray lines).

an inner dimension of 4.5×3.9×2.7 m. A wire-cable meshwork formed a suspension floor over the floor cones. The Metadyne cones and their supporting panels (IAC Noiselock II) have an absorption coefficient of 0.99 between 0.3 and 8 kHz. Noise was reduced by at least 65 dB above 0.5 kHz and by 85 dB in the frequency range relevant to spatial hearing in the barn owl (3–9 kHz). The chamber deviated from the inverse-square law by no more than ±1.5 dB between 3 and 10 kHz measured along transects extending between opposite corners of the room. The ambient noise inside the chamber was below 18 dB SPL_A.

1. Acoustic stimulation

Stimuli were presented from three speakers mounted on a hoop forming an arc in the azimuthal plane at eye level (Fig. 1). For measuring discrimination of sources at 3° separations, speakers were positioned 1.98 m from the owl's head. For all other measurements, speakers were positioned at a distance of 0.79 m. Stimulus waveforms were generated digitally on a computer and converted to analog form at 30 000 samples per second (Tucker-Davis Technologies Power DAC PD1 or Modular Instruments M308). Analog signals were attenuated with programmable attenuators (PA4, Tucker Davis Technologies, or M300, Modular Instruments), amplified (HB6, Tucker-Davis Technologies, or MacIntosh 754) and delivered to the speakers (Peerless, 2-in. cone tweeters). Stimuli were ''frozen'' noise segments (i.e., the same waveform was repeated on each trial) with durations of 25 ms, for owls 889 and 882, or 100 ms, for owl 896, with onsets and offsets shaped by 2.5-ms cosine ramps. Frozen noise was used to prevent false positive responses that might otherwise occur if the owls could detect the differences between random noise stimuli (see *Behavioral measurements*, below). Digital stimulus waveforms were generated with flat spectra from 3 to 9 kHz. The final stimulus spectrum, shaped by the transfer functions of the speakers, was flat from 3 to 9 kHz (±3 dB). Transfer functions of the different speakers did not differ by more than 2.5 dB at any frequency from 3 to 9 kHz. Stimulus sound-pressure level was 74 dB SPL_A, measured with a $\frac{1}{2}$-in. microphone (Brüel & Kjaer model 1760) and sound-level meter (Brüel & Kjaer type 2235) at the approximate position of the owl's head before the beginning of each test session.

2. Behavioral measurements

The acoustically evoked PDR was measured using an infrared (IR) pupillometer, consisting of an IR-emitting diode and an IR-detecting diode [Figs. 1, 2(A)]. The right eye was held open by threads taped to the eyelids, leaving the nictitating membrane free to moisten the cornea. The pupillometer was positioned such that the IR emitter illuminated the entire cornea, and the detector sampled light reflected by the ventral-nasal quadrant of the retina. The voltage output of the IR detector, thus, varied in proportion to the dilation of the pupil. Output of the IR detector was amplified, digitized at a sampling frequency of 1875 samples per second, and stored on computer disk. The magnitude of the PDR was quantified by summing the area under the IR-detector voltage trace for 2 s following stimulus onset, after first ''zeroing'' the trace by subtracting the voltage level measured at sound onset.

Spatial discrimination was measured using a method based on habituation and recovery of the PDR described in a previous report (Bala and Takahashi, 2000). The PDR is a reflexive dilation of the pupil evoked by any salient, novel stimulus, such as a sound, in the owl's environment. Repetition of the same sound causes the PDR to habituate. After the

PDR has habituated to one sound, presentation of a different sound can evoke a ''recovered'' PDR, with amplitude comparable to that of the response to initial presentation of the first sound. Thus, the degree to which a second *test* sound recovers the PDR from habituation to the first *habituating* sound can be used as a measure of the owl's ability to detect the difference between the two sounds.

Individual test sessions consisted of 300 stimulus trials with an intertrial interval varying randomly from 9 to 12 s. In a single session, a constant habituating stimulus was presented for the first 119 trials. Following this initial habituation period, a test stimulus was presented every 40 trials. Spatial discrimination was measured in three conditions (Fig. 1): (1) In the *single source* condition, the habituating stimulus consisted of a noise burst presented from a speaker at 6° azimuth (Fig. 1, ''H''). On test trials the same noise burst was presented from another speaker, displaced to the right (Fig. 1, ''T''). The azimuthal separations tested included all (owl 882) or some (owls 889 and 896) of the following: 3°, 5°, 7°, 10°, 15°, 20°. (2) In the *lead source* condition, the habituating stimulus consisted of a noise burst from the speaker at 6° azimuth followed, at a delay of 3 ms, by the same noise burst from a speaker at a fixed location of −20° azimuth (Fig. 1, ''F''). On test trials the azimuth of the lead source was changed. Each owl was tested with all (owl 882) or some (owls 889 and 896) of the following lead source separations: 3°, 5°, 10°, 15°, 20°, 25°. (3) In the *lag source* condition, the habituating stimulus consisted of noise burst from the speaker at −20° azimuth followed, at a delay of 3 ms, by the same noise from the speaker at 6° azimuth. On test trials the azimuth of the lag source was varied. Each owl was tested with a subset of the following lag source separations: 3°, 5°, 10°, 15°, 20°, 25°, 30°, 35°, 40°. The speakers at positions H, T, and F were always the same, regardless of test separation, to control for the possibility that the owls could distinguish between habituating and test speakers. Using the same frozen noise segment as the stimulus on all trials insured that recovered responses were not evoked by detectable differences in the stimulus waveforms across trials, other than those caused by changing the location of a speaker. Discrimination performance was quantified using the nonparametric statistic, proportion correct $[p(c)]$, calculated from empirical receiver operating characteristic (ROC) curves, as described in the Results section.

Minimum audible angles were estimated from psychometric functions obtained by plotting discrimination performance $[p(c)]$ as a function of azimuthal separation between habituating and test sources. A Gauss–Newton least-squares method was used to fit the psychometric data with sigmoid functions of the form

$$p(c) = \delta - (\delta - 0.5)\exp[-(x/\alpha)^{\beta}], \quad (1)$$

where x is azimuthal separation between habituating and test speakers, α is a threshold parameter that determines the displacement of the function along the abscissa, β is a slope parameter (Britten *et al.*, 1992). The parameter corresponding to asymptotic performance level at large azimuthal separations, δ, was made equal to the highest performance level measured for each owl under each test condition. This constraint resulted in more meaningful sigmoid fits for lag discrimination by owls 882 and 896, and otherwise had little effect on the resulting curve fits, particularly within the region of steepest slope that was used for threshold determination. This method resulted in significant fits (F test, $p < 0.01$) for all conditions in all owls, with one exception (owl 896, single sources). In cases where the curve fit was significant, the MAA was estimated from the fit curve by determining the source separation corresponding to a discrimination performance level of $0.71\,p(c)$. The single source MAA for owl 896 was conservatively estimated as the smallest source separation for which discrimination performance exceeded $0.71\,p(c)$.

B. Results

Spatial discrimination for single, lead, and lag sources was measured using habituation and recovery of the PDR in three owls. Representative habituation and recovery sequences from a single owl (owl 882) are illustrated in Figs. 2 and 3. Initial presentation of the habituating stimulus (consisting of a 25-ms noise burst from the speaker at 6° azimuth, followed at a 3-ms delay by the same noise from the speaker at −20° azimuth) evokes a large pupillary dilation, as indicated by a large positive voltage deflection [h1 in Fig. 2(B)]. On subsequent trials (h2 through 4), the response to repetitions of the same stimulus diminishes progressively. After the magnitude of responses to the habituating stimulus has declined to a relatively stable level, a test stimulus is presented. In this example, the location of the lead sound is changed by 20°. The magnitude of the PDR evoked on this test trial [Fig. 2(C), t1] is larger than those evoked on the preceding habituating trials (gray lines). Thus, a change in lead source location serves to recover the PDR from habituation.

Responses of the same owl from four test sessions with the same stimulus configuration show similar patterns of habituation and recovery (Fig. 3). In each session, the amplitude of responses to the habituating stimulus exhibits a decreasing trend that is most evident for the first ten trials, and continues until a relatively stable response level is obtained by trial 80, or sooner. The magnitude of responses to test stimuli (unfilled large circles) is generally larger than that to habituating stimuli (filled circles). Thus, both habituation and recovery from habituation appear to be stable. To facilitate comparison of our results with the majority of previous human and animal studies of the precedence effect, an attempt was made to use shorter stimuli. However, in initial experiments measuring single source discrimination with sound durations of 15 ms or less it was observed that, following the first test stimulus, responses to successive test stimuli habituated. For owls 889 and 882, stable recovery from habituation was obtained when stimulus duration was increased to 25 ms. For owl 896, stimulus duration had to be increased to 100 ms to achieve stable recovery.

In the test sessions shown, it is clear that the mean response magnitude for test stimuli is greater than that for habituating stimuli, but there is also considerable overlap between the distributions of response magnitudes for the two types of stimuli. The difference in mean response magnitude

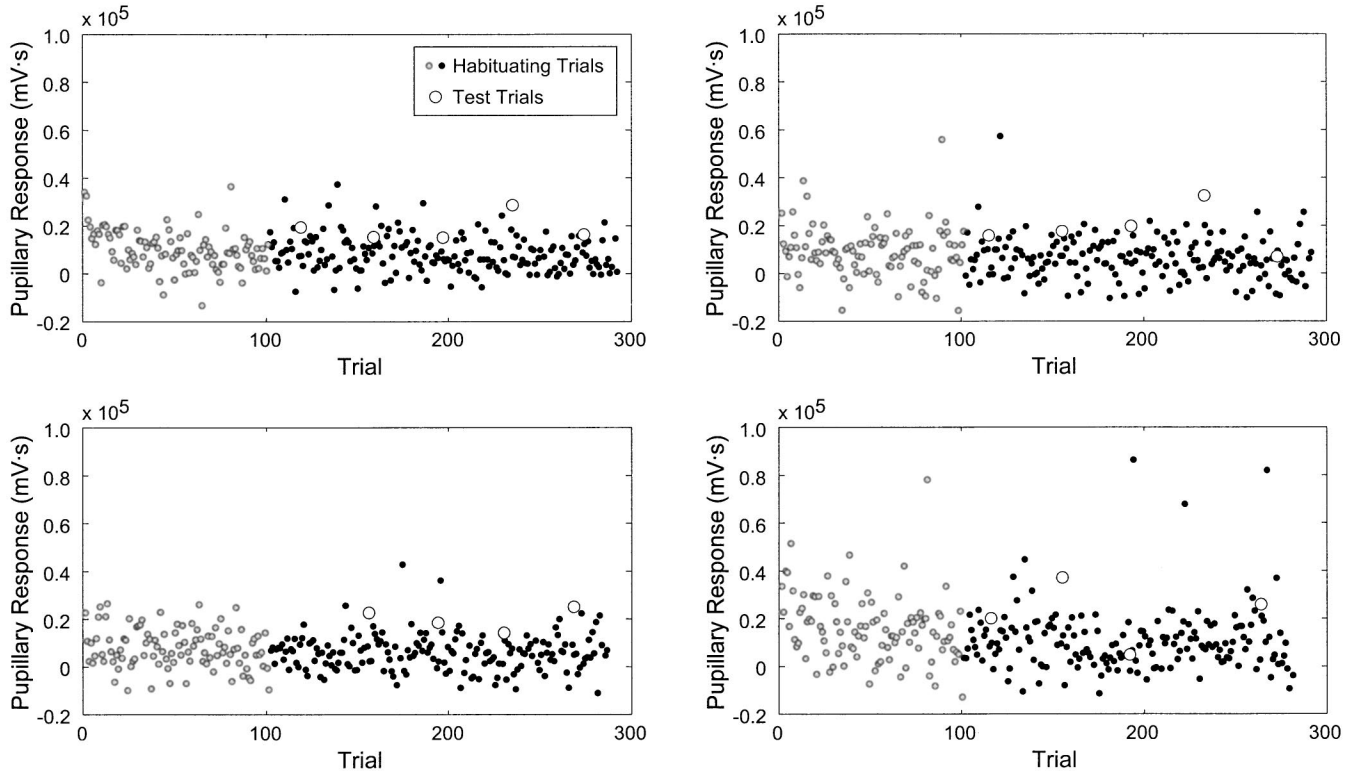


FIG. 3. Consistent patterns of habituation and recovery are observed on four test sessions with the same owl (882) and test stimuli. The pupillary response magnitude on each trial is measured by summing the area under the pupillometer voltage trace during the first 2 seconds following stimulus onset. Stimulus configuration is the same as in Fig. 2. Responses to the first 100 habituating trials (gray circles) are eliminated prior to analysis of discrimination performance. The total number of recorded responses is less than 300 because trials contaminated by blinks were eliminated.

for test and habituating stimuli, relative to the pooled variance of responses, can provide a measure of the owl's ability to discriminate the test stimulus from the habituating stimulus. Because the number of test stimuli in a single session had to be limited to prevent the owl from habituating to the *test* stimulus, responses were pooled across multiple sessions to obtain reliable estimates of discriminability for a given test condition. Despite the overall similarity of responses, however, there were differences in mean response magnitude and variance between sessions, most likely reflecting small differences in placement of the pupillometer relative to the eye, and in retraction of the eyelids. In order to pool responses across test sessions it was necessary to first normalize responses from individual sessions. For this purpose, the amplitudes of responses to individual habituating and test trials were converted to z-scores, relative to the distribution of habituating trial responses after trial 100. Responses from the initial 100 trials of each session (Fig. 3, gray dots) were discarded to insure that the calculated mean reflected the amplitude of the fully habituated PDR.

The owls' abilities to discriminate between test and habituating stimuli were quantified by the nonparametric statistic, $p(c)$, calculated by determining the area under probability-space ROC curves generated from the pupillary responses. A nonparametric discrimination measure was used because the variance of responses to test and habituating stimuli often differed. To generate ROC curves, normalized responses were pooled from four or more test sessions with the same test condition and test azimuth separation from a single owl. ROC curves were generated by applying methods developed by Britten *et al.* (1992) for use with neuronal response data. The continuously distributed PDR amplitude data were treated as results of a rating experiment, in which the amplitude of the response on each trial was considered as the subject's rating of the likelihood that a test stimulus was presented. Following this assumption, ''hits'' were defined as responses to test stimuli that exceeded an arbitrary response criterion, and ''false alarms'' were defined as responses to habituating stimuli that exceeded criterion. Probability-space ROC curves were then obtained by plotting the proportion of hits against the proportion of false alarms for each of a set of criterion values, spanning the range of measured response amplitudes, and including values above and below the extreme response amplitudes. The discrimination metric, $p(c)$, is calculated by determining the area under the ROC curve. This quantity varies from 0.5, indicating chance performance, to 1.0, indicating perfect discrimination, and is equivalent to the proportion correct in a two-alternative forced choice task (Green and Swets, 1966).

Examples of ROC curves obtained from pooled response data for single-, lead-, and lag-source discrimination by owl 889 are shown in Fig. 4. In each test condition, the ROC curve for the smallest test separation is close to the unity line (dashed), indicating discrimination performance near chance levels $[p(c)=0.5]$. The area under the ROC curves generally increases as test separation is increased, reflecting an increase in the mean amplitude of test responses relative to habituated responses. Note, however, that in the single- and lead-source conditions, maximum discrimination was achieved at the second largest test separation. In a few cases,

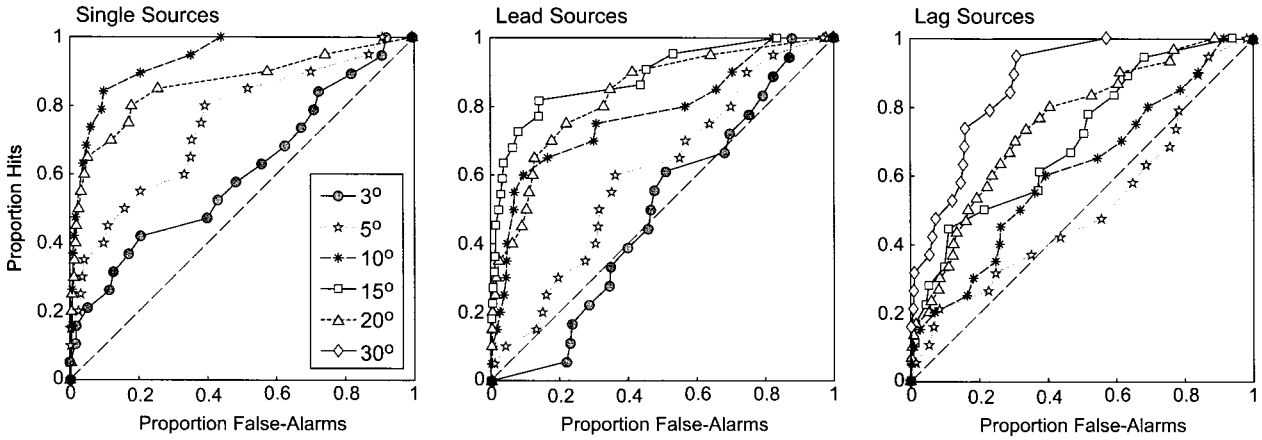


FIG. 4. ROC curves generated from pupillary responses of owl 889 in three test conditions. Individual curves were generated by plotting the proportion of responses to test and habituating trials that exceeded criterion (''hits'' and ''false alarms,'' respectively), for each of a set of criterion amplitude values (see the text for details). Each plot shows ROC curves generated for each azimuthal separation tested in a single test condition. Key shows symbols for all three plots. Discrimination performance for each test separation was quantified by calculating the area under the corresponding ROC curve to give the nonparametric statistic $p(c)$.

the ROC curves for discrimination of large test separations were clearly asymmetric (e.g., single source: 20°; lead source: 10°, 15°). This effect occurs when the variance of responses to test and habituating responses differs (Green and Swets, 1966), and was often observed at larger test separations because the variance of test response amplitudes tended to increase in proportion to the mean amplitude.

Psychometric functions for azimuthal discrimination of single, lead, and lag sources were obtained from three owls by plotting $p(c)$ as a function of test separation (Fig. 5). For all three owls, the discrimination functions for lag sources (squares) have shallower slopes than those for single (circles) and lead sources (triangles). Consequently, for arbitrary criterion performance levels from 0.65 to 0.85 $p(c)$, the MAAs for lag sources are greater than those for lead sources, with the magnitude of threshold elevation being dependent on criterion level. In addition, for all three owls, the MAA for single sources was less than that for lead sources. Thus, the ability to resolve sound source locations in either of the two-source conditions was impaired relative to that in the single-source condition, with the greater impairment occurring in the lag-source condition. Discrimination performance of each owl reached chance levels at the smallest test separations in the lead- and lag-source conditions, indicating that owls were unable to discriminate between the speakers used to present habituating and test stimuli.

Discrimination performance of each owl was summarized by determining MAAs for the three test conditions (Table I). MAAs were estimated by fitting the discrimination data to sigmoid curves (see Methods), and calculating the angular separation at which the fit curves achieved an arbitrary threshold value of 0.71 $p(c)$. This threshold measure was used, in preference to the threshold parameters of the curve fits, because the individual psychometric functions for single- and lead source conditions appear to reach saturation at different $p(c)$ values, and because we did not test sufficiently large lag source separations to be sure that the lag source performance curves reached saturation. The 0.71 $p(c)$ criterion was chosen to be consistent with the most similar human psychophysical study (Litovsky, 1997).

An alternative, statistical criterion was also used to measure azimuthal discrimination thresholds. The statistical threshold (ST) is defined as the smallest test separation at which the mean amplitude of responses to test stimuli was significantly greater than the mean amplitude of responses to habituating stimuli ($p<0.005$, Mann-Whitney U-test). Responses to test and habituating stimuli were pooled across test sessions as in the preceding analysis. ST values obtained in this manner are shown in Table I. Within the limits of sampling resolution, results obtained with this method agree with those of the former analysis: In all cases, both methods identify transitions from subthreshold to suprathreshold performance between the same pairs of test separations.

Threshold elevation factors (TEF), defined as the ratio of MAAs in the two-source and single-source conditions (Litovsky, 1997), are also shown in Table I to facilitate comparison with results of previous human psychophysical studies. A significant curve fit could not be obtained for owl 896 in the single source condition. In this case, the MAA was estimated as the smallest azimuthal separation (3°) at which discrimination performance exceeded threshold. This value is in agreement with results of a previous study, using similar methods, in which it was shown that the owl's MAA measured using 100-ms sounds is approximately 3° (Bala and Takahashi, 2000)

III. EXPERIMENT 2: DISCRIMINATION OF NONSPATIAL CUES ASSOCIATED WITH LEAD–LAG COMBINATIONS

Results of the preceding experiment suggest that sensitivity to the binaural cues that barn owls use to localize sounds (Moiseff, 1989) is reduced, but not eliminated, in simulated echoic conditions. This conclusion depends on an assumption that discrimination was based on the difference in perceived location of habituating and test stimuli. It is possible, however, that discrimination was based on differences in the perceived quality of sounds between test and

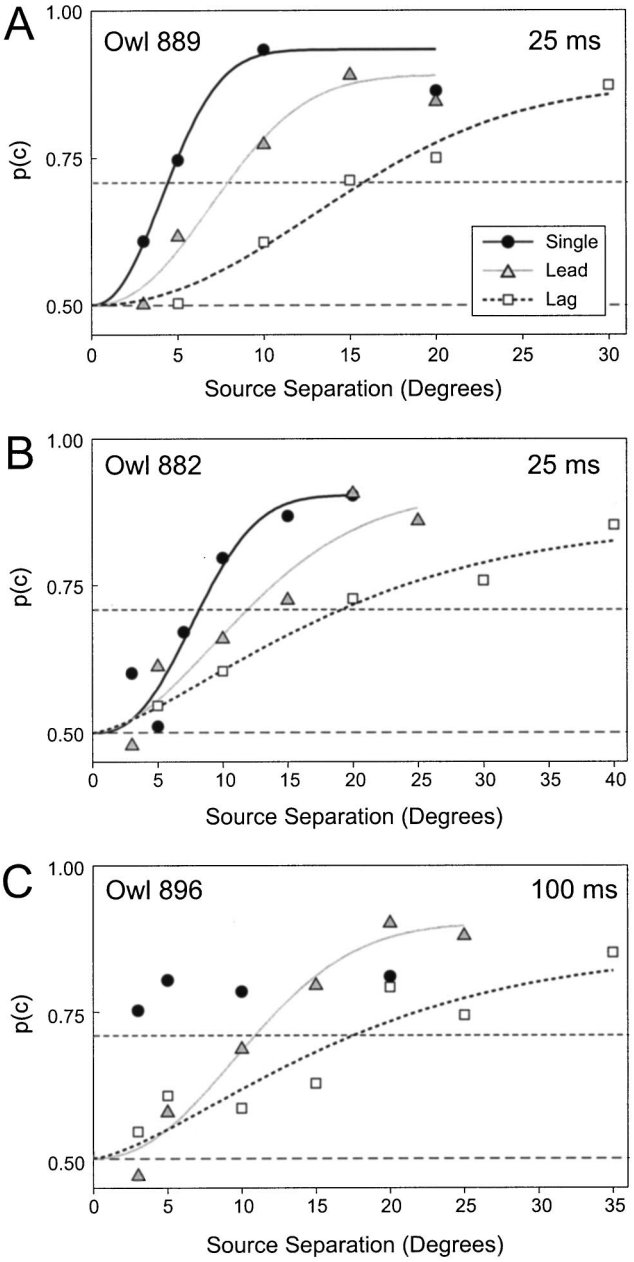


FIG. 5. Psychometric functions for azimuthal discrimination of single, lead, and lag sources by three owls. Each point represents the proportion correct $[p(c)]$, calculated as the area under the probability space ROC curve obtained from pooled responses from a minimum of four test sessions, and including a minimum of 18 test responses. The curve fit for owl 896, single source condition was not significant ($p<0.05$), and is therefore omitted. The 0.71 $p(c)$ criterion level used to calculate MAAs is indicated by a grid line in each plot.

habituating trials. The stimulus configuration used in experiment 1 gives rise to spectral and temporal cues at either ear that could provide a basis for discrimination. These cues are illustrated in Fig. 6.

The first cue is the rippling of spectra that results when the lead and lag waveforms sum in each ear. The period of the ripple changes as a function of the location of the lead and lag sources. Figure 6(A) illustrates the difference in the ripple period in the right ear when the lag source was at 6° (solid) and at 36° (dotted line). The spectra shown in Fig. 6(A) were generated by addition of lead and lag waveforms that were filtered using an owl's head-related transfer functions for the appropriate lead (−20°) and lag (6° or 36°) locations. Lead–lag delay was 3 ms. The period of spectral rippling, equal to the reciprocal of the delay between lead and lag waveforms, is plotted as a function of lag source azimuth in Fig. 6(B) for habituating and test stimuli used in the lag discrimination condition of experiment 1. Changing the azimuth of the lag source results in orderly changes of the ripple period at the left (circles, solid line) and right (×'s, dotted line) ears. In humans, changes in spectral ripple period give rise to changes in perceived pitch (Yost *et al.*, 1978). In addition to spectral rippling, changing the location of one source causes small differences in the overall shape of the power spectrum [i.e., differences in heights of neighboring peaks in Fig. 6(A)] and changes in the temporal fine structure of monaural signals. The latter effect is illustrated in Fig. 6(C). For each ear, the waveform received in the habituating condition was cross correlated with that received in each of the lag-source test conditions. The dotted and solid functions plot the maxima of these normalized cross correlograms (coherence of monaural signals) against the lag source's azimuth for the right (dotted) and left (solid) ears, respectively. Increasing the azimuthal separation of lag sources on habituating and test trials results in a orderly decrease in coherence of the monaural signals at either ear, providing additional potential cues for discrimination.

It is thus possible that the owls discriminated test trials from habituating trials by detecting changes in the quality of the sounds and not changes of their perceived locations. If this were the case, results of experiment 1 may underestimate the reduction of spatial sensitivity in echoic conditions. To control for this possibility, we measured discrimination using stimuli approximating the signals that would have been present at one ear on habituating and test trials of experiment 1, presented from a single, stationary speaker.

A. Methods

Experimental subjects and PDR measurement procedures were the same as those in experiment 1. General aspects of experimental design were also similar, so only differences will be detailed. All stimuli were presented from a single speaker at 6° azimuth. The initial habituation period consisted of 99 trials. Three different test stimuli, corresponding to different lag-source locations, were tested in each session. After 100 habituating trials, test stimuli were presented in blocks of three, with each test trial separated by nine intervening habituating trials. Test blocks were separated by 20 habituating trials. Five test blocks were presented in each session.

Stimuli similar to the waveforms resulting from the summation of lead and lag sounds at an owl's ear in experiment 1 were generated by filtering sounds with head-related transfer functions (HRTFs) corresponding to lead and lag source locations, delaying the lag signal by 3 ms, and adding the resulting waveforms. Filtering with HRTFs replicates the transformation of both phase and amplitude spectra that occurs between a source at a given location and a position in the external ear canal immediately adjacent to the tympanum. HRTFs were measured as previously described (Keller

TABLE I. Minimum audible angles (MAA) and statistical thresholds (ST) for the three owl subjects. MAA values represent the angular separation corresponding to a discrimination performance of $0.71\,p(c)$, obtained from the sigmoid curve fits and rounded to the nearest degree. ST values represent the smallest tested azimuthal separation at which the mean amplitude of responses to test stimuli, pooled across sessions, was significantly greater than the mean amplitude of responses to habituating stimuli. Threshold elevation factors (TEF), shown for the two-source conditions, are the ratio of the corresponding MAA to the single-source MAA. The single-source MAA and associated TEF values for owl 896 are shown in parentheses to indicate that they are estimated values (see the text).

	Condition							
	Single source		Lead source			Lag source		
Owl ID (sound duration)	MAA	ST	MAA	ST	TEF	MAA	ST	TEF
889 (25 ms)	4°	5°	8°	10°	2	16°	20°	4
882 (25 ms)	8°	10°	12°	15°	1.5	19°	20°	2.4
896 (100 ms)	(3°)	3°	11°	15°	(3.7)	18°	20°	(6)

et al., 1998) in an owl not used as a test subject. Using this method, we measured the ability of owls to discriminate stimuli that approximate the waveforms present at one ear on habituating and test trials in experiment 1. Stimuli corresponding to those in the lag discrimination condition were chosen, because spatial discrimination was most impaired for lag sources.

The input stimulus waveforms were the same frozen noise segments used in experiment 1. To generate the lead sound, the waveform was convolved with the time-domain representation of the HRTF for the lead location. To generate the lag sound, a 3-ms onset segment, consisting of zeros, was concatenated to the beginning of the noise, and the resulting signal was convolved with the HRIR for the lag location. The final signals, approximating the monaural signals for each pair of lead and lag locations, were generated by adding the appropriate pairs of lead and lag signals. Because the HRTFs used were measured in another owl, the resulting stimuli do not exactly replicate the monaural waveforms experienced in experiment 1. However, the relative time delays for sources at different locations, that give rise to spectral rippling, are very similar in different owls (Keller *et al.*, 1998). Furthermore, prominent spectral notches that give rise to large individual differences in HRTFs of mammals (Middlebrooks *et al.*, 1989; Musicant *et al.*, 1990; Spezio *et al.*, 2000) are not present in the HRTFs of barn owls at the locations along the frontal horizon used in this study (Keller *et al.*, 1998). Thus, although the method used does not exactly replicate the monaural signals experienced in experiment 1, it generates habituating and test signal waveforms that differ in spectral shape and temporal correlation in a manner similar to the actual monaural waveforms in that experiment.

Owls were tested for their ability to discriminate the different monaural signals generated in the lag source discrimination condition of experiment 1. Thus, in each case the habituating stimulus was generated by filtering the lead waveform with the left- or right-ear HRTF for $-20°$ azimuth, and the lag waveform with the same ear's HRTF for $+6°$ azimuth. The test stimuli were generated by filtering the lead waveform with the HRTF for $-20°$, and the lag waveform with the HRTFs corresponding to a subset of the lag source locations tested in experiment 1. In this manner the three owls were tested for discrimination of monaural combination equivalent to the following lag source test separations:

Owl 889...left ear: 5°, 10°, 15°, 20°, 30°
Owl 882...left ear: 10°, 20°, 30°; right ear: 5°, 10°, 20°, 30°, 34°
Owl 896...left ear: 10°, 20°, 34°; right ear: 3°, 15°, 34°.

B. Results

Discrimination performance for the summed lead–lag signals is plotted as a function of lag source azimuth separation for test stimuli in Fig. 7. Discrimination performance for left- and right-ear signals is plotted as black and gray circles, respectively. The psychometric curves for lag source discrimination from Fig. 5 are replotted for comparison (squares, dotted lines). Owls 882 and 896 were unable to discriminate any of the test signals from the habituating signal [Figs. 7(B) and (C)]. Owl 889 achieved a maximum discrimination level of $0.69\,p(c)$ for the combination signal corresponding to a lag source separation of 10° [Fig. 7(A)]. Discrimination performance declined to chance levels at larger equivalent source separations. There was no indication that discrimination performance increased with increasing equivalent source separation for any of the subjects, as would be expected if lag-source discrimination were dependent on sensitivity to monaural cues. Furthermore, in all cases discrimination performance was at chance levels for combination signals corresponding to lag source separations that were themselves discriminable in experiment 1. The preceding results were confirmed by statistical analysis of responses, pooled across sessions. For owl 889, the mean response to the test stimulus equivalent to a 10° separation was significantly greater in amplitude than the mean response to the corresponding habituating trials at a significance level of $p=0.007$ (Mann-Whitney U test). For no other stimulus in any owl was there a significant ($p<0.05$) difference between response amplitudes on test and habituating trials. These results indicate that sensitivity to the spectral and temporal cues available at either ear is insufficient to account for discrimination of changes in lag source location.

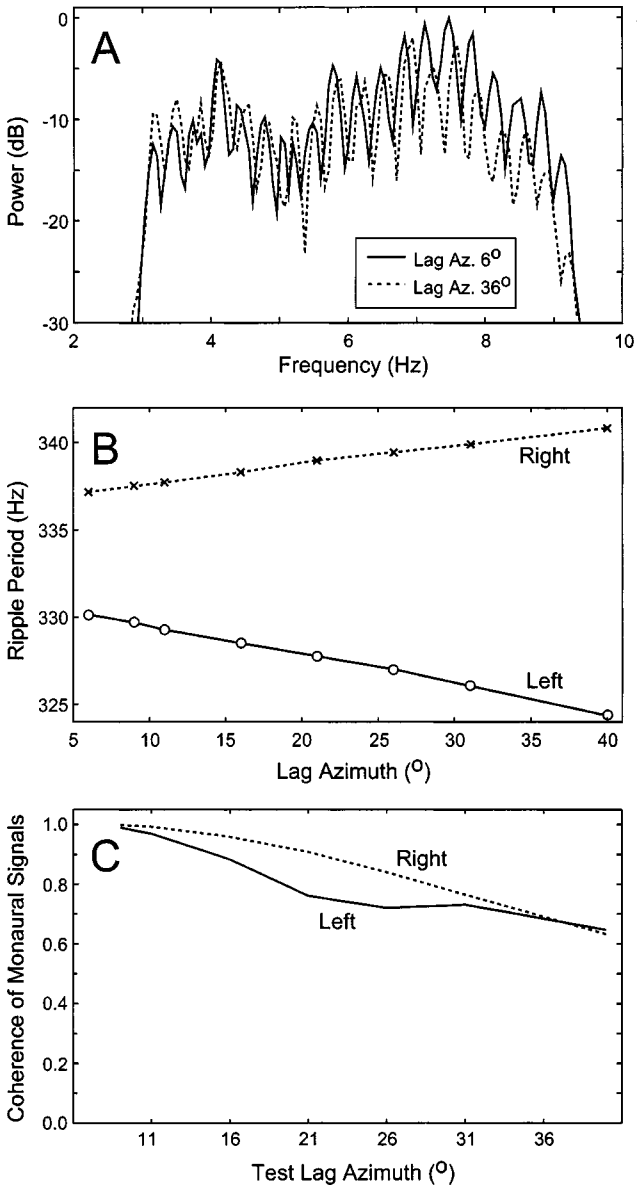


FIG. 6. Monaural cues to lag source location in experiment 1. (A) Power spectra of the signals at an owl's left ear generated by lead–lag pairs that differ only in location of the lag source. Addition of lead and lag waveforms filtered by an owl's HRTFs simulates the monaural signals generated in the lag-source condition of experiment 1 by the habituating (6° lag azimuth) and 30° test stimuli. (B) Ripple period of the monaural signals changes linearly as a function of lag-source azimuth. As the lag source is moved to the right (positive azimuths) the ripple period increases at the right ear (dotted line) and decreases at the left ear (solid line). (C) Coherence of waveforms at the left (solid line, circles) or right ears (dotted line, ×'s) on habituating and test trials decreases monotonically with increasing lateral displacement of the lag source. Coherence is defined as the maximum of the normalized cross-correlation function calculated for the monaural signals present on habituating and test trials.

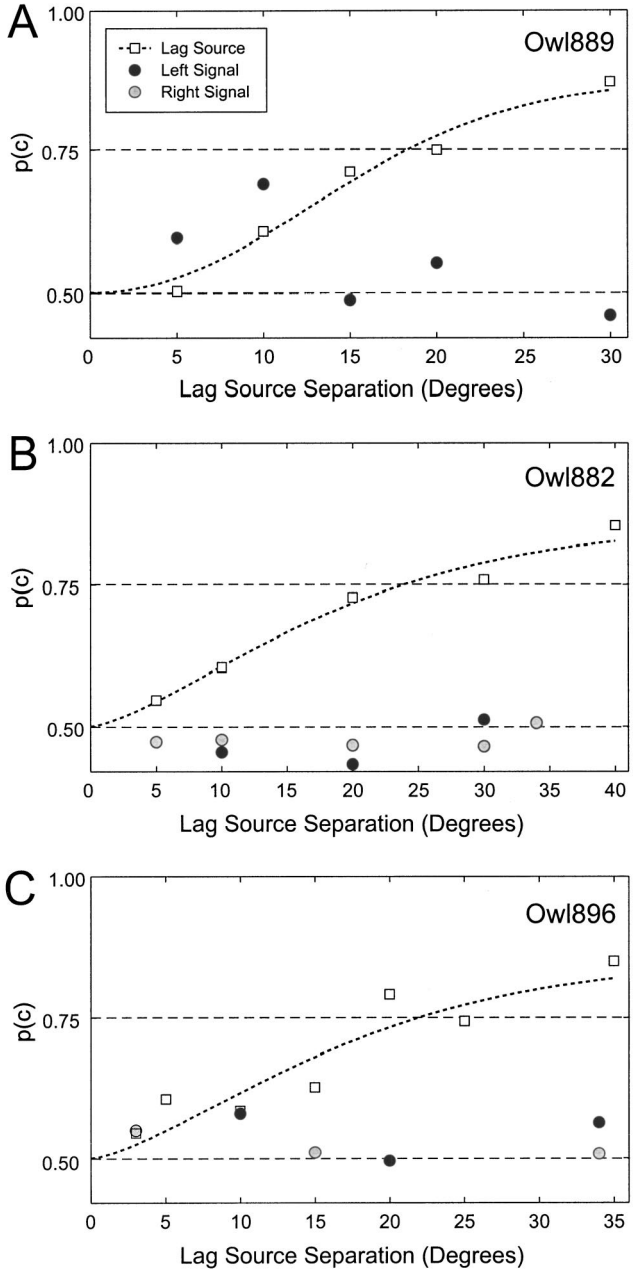


FIG. 7. Psychometric functions for discrimination of lead–lag signal combinations presented from a single speaker, and for equivalent lag sources for three owls (A–C). Discrimination of lead–lag combination signals generated using the left (black circle) and right (gray circle) ear HRIRs is plotted as a function of the equivalent lag-source separation, defined as the difference in azimuthal location of the HRIRs used to generate the lag-source signal component of habituating and test stimuli. Duration of lead and lag sounds was 25 ms for owls 889 and 882, and 100 ms for owl 896. Each point represents the proportion calculated from pooled responses from a minimum of four test sessions, and a minimum of 18 test responses. Lag-source discrimination data (squares) are replotted from Fig. 5.

IV. DISCUSSION

This study represents the first application of a spatial discrimination paradigm to measure the precedence effect in an animal model. The results demonstrate that the ability to detect a change in the location of a sound source is impaired in the presence of a simulated reflection. The ability to localize the simulated reflection itself is even further impaired. Results of the control experiment (experiment 2) demonstrate that discrimination of changes in source location in echoic conditions was not based on sensitivity to changes in sound quality associated with changes in the monaural signals. Thus, the impairment of spatial resolution in simulated echoic conditions most likely reflects a reduced ability to use binaural directional information.

Results for azimuthal discrimination obtained in experiment 1 are similar to those obtained in comparable human psychophysical studies. Several previous human studies have

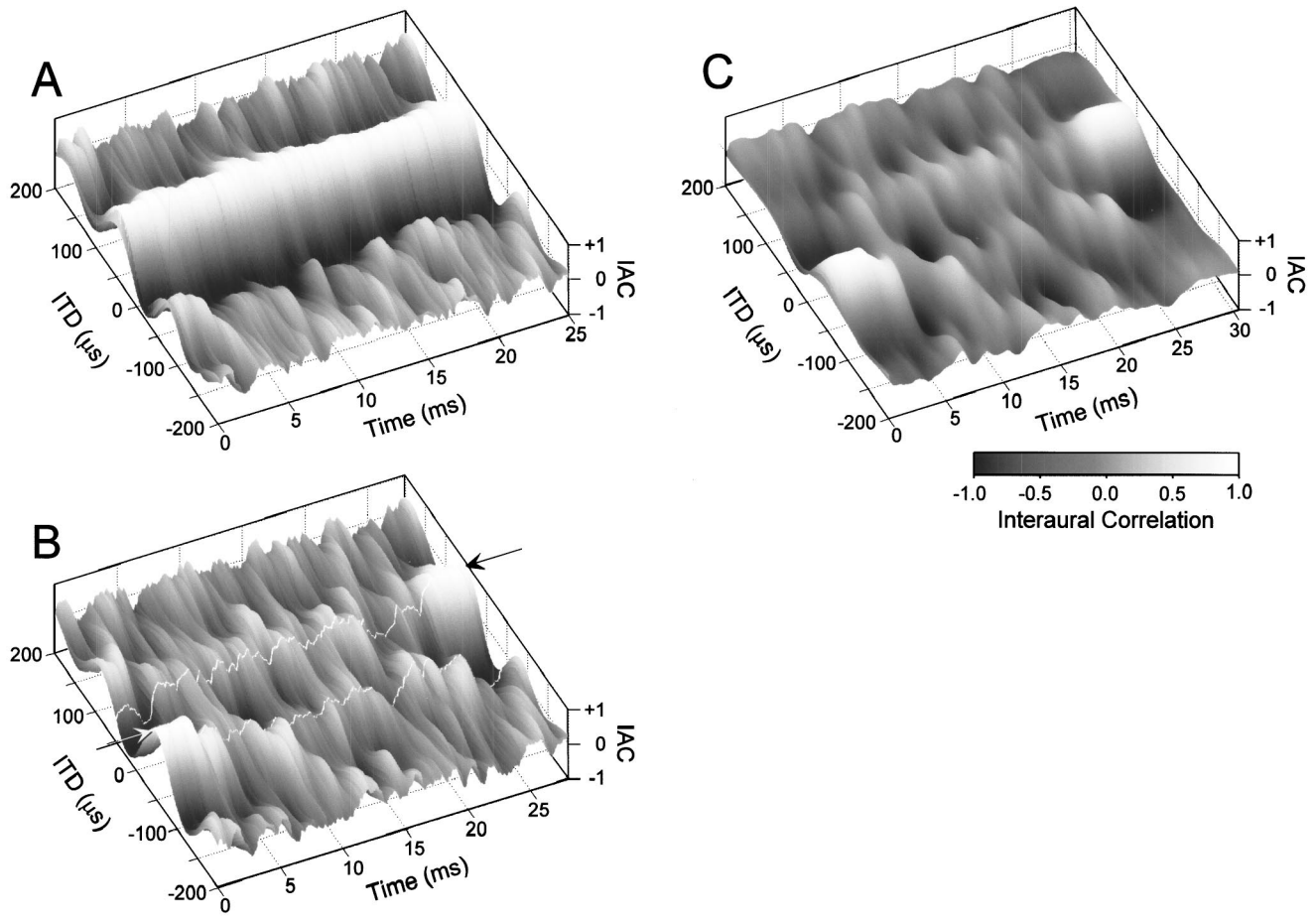


FIG. 8. Running interaural cross correlograms illustrate the effects of lead–lag superposition on ITD cues available in experiment 1. (A) A single source at +20° azimuth generates a constant ITD of 49 μs. In this and the following plots the monaural acoustic signals were simulated by filtering the stimulus waveform used in experiment 1 with an owl's head-related transfer functions. Running interaural cross correlograms were constructed by calculating the cross-correlation function of the signals at the two ears in a 0.5-ms window that was advanced in 33-μs increments throughout the total stimulus duration. Running time is plotted on the *x* axis, ITD on the *y* axis (positive values indicate the right signal leads), and interaural correlation (IAC) on the *z* axis and color scale. (B) The combination of a lead source at −20° azimuth and a lag source at +20°, with a lag delay of 3 ms, generates a time-varying pattern of interaural correlation. The ITDs of lead and lag sources, presented in isolation, are indicated by pointers and white lines on the plot surface. Here, the monaural signals were generated by summation of lead and lag waveforms that had been filtered by the HRTFs for the appropriate source locations. (C) The effects of peripheral filter response times on the available ITD cues are illustrated by constructing the interaural cross correlogram after passing the monaural signals through a bank of gamma-tone filters (Slaney, 1993), modified to match the bandwidths of barn owl auditory-nerve fibers (Köppl, 1997). At each time increment, the interaural cross-correlation function was computed separately in each frequency channel. The resulting frequency-specific cross-correlation functions were then averaged and displayed as in plot B.

demonstrated lag discrimination suppression (Litovsky *et al.*, 1999), an impaired ability to detect changes in the azimuthal location or binaural attributes of lagging sounds relative to either single- or leading-sounds. This effect has been demonstrated using either free-field (Perrott *et al.*, 1989; Litovsky and Macmillan, 1994; Litovsky, 1997) or headphone (Zurek, 1980; Saberi and Perrott, 1990; Shinn-Cunningham *et al.*, 1993; Tollin and Henning, 1998; Litovsky and Shinn-Cunningham, 2001) stimulus presentation. The human study that used the most similar stimulus presentation methods and parameters to those employed in the present study also reported lead discrimination suppression (Litovsky, 1997). Furthermore, the magnitude of impairments for lead and lag sources reported in that study, as quantified by TEFs, are roughly equivalent to those measured with stimuli of the same duration (25 ms) in the present study. The similarity of the present results to those of previous human studies suggests that similar processes may give rise to the precedence effect in barn owls and humans.

A. Effect of acoustic interactions on available binaural cues

Azimuthal sound localization by barn owls is primarily dependent on sensitivity to interaural time differences (ITD; Moiseff, 1989; Poganiatz *et al.*, 2001), which are detected via neural processes analogous to interaural cross correlation (Carr and Konishi, 1990). One factor that may contribute to impaired spatial resolution in echoic conditions is degradation of ITD cues caused by acoustic superposition of lead and lag sounds. The effect of superposition of lead and lag waveforms on the ITD cues available in experiment 1 is illustrated in Fig. 8. A single source at +20° azimuth generates a constant ITD of 49 μs, as indicated by the single peak of the running interaural correlogram in Fig. 8(A). This and the following plots were generated by computing the cross-correlation functions of acoustic signals at the left and right ears in a 1-ms time window that was advanced, in 33-μs increments, throughout the total stimulus duration. This time

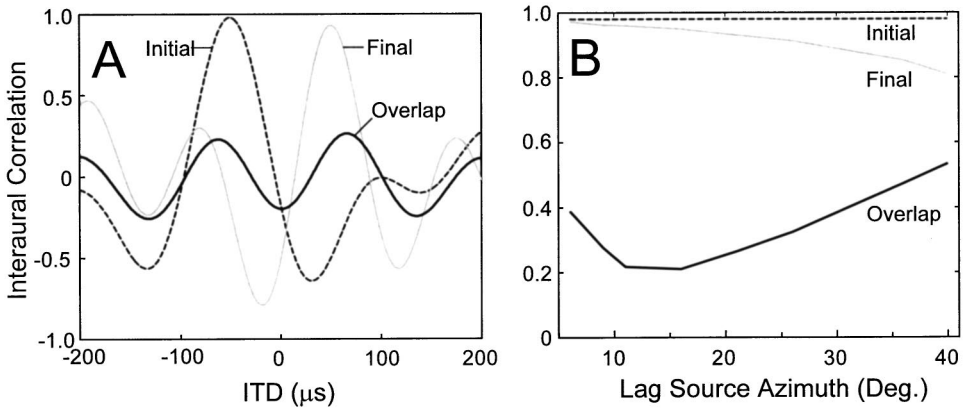


FIG. 9. Interaural correlation values computed over the entire durations of the overlapped and nonoverlapped stimulus segments. (A) Normalized interaural correlation functions were computed in rectangular time windows encompassing the 3-ms initial (dashed) and final (solid gray) stimulus segments and the 22-ms overlap segment (solid black). Stimulus parameters are the same as in Fig. 8(A). (B) Maximum interaural correlation levels in the three stimulus segments are plotted as a function of lag-source azimuth for stimulus configurations used in the lag-source discrimination condition of experiment 1.

window duration was chosen to illustrate the available ITD cues with a temporal resolution exceeding that of the barn owl's binaural system (Wagner, 1991), while being long enough to allow the lowest frequency stimulus components to contribute to the cross-correlation function. When a leading source is added at −20° azimuth, with a lead–lag delay of 3 ms, the pattern of interaural correlation over time becomes more complex [Fig. 8(B)]. In the initial and final 3-ms stimulus segments, only one of the two sources is active, and prominent peaks of interaural correlation occur at ITDs of −49 or +49 μs, corresponding to the lead or lag source locations, respectively. These ITDs are indicated by pointers and white lines on the plot surface. In the intervening 22-ms segment, both sources are active and superposition of lead and lag waveforms at each ear results in a complex, time-varying pattern of interaural correlation. Throughout much of this overlap segment, peaks are still evident at, or near ITDs of −49 and +49 μs. However, the level of interaural correlation at each of these ITDs varies considerably over time, and peaks of interaural correlation also occur, intermittently, at other ITDs. The average pattern of interaural correlation within the initial, overlap, and final stimulus segments is illustrated in Fig. 9(A). Within the overlap segment, two peaks are evident near the ITDs of lead and lag sources. However, the maximum level of interaural correlation is substantially reduced relative to the initial and final segments. The peak level of interaural correlation within the overlap stimulus segment varied with test speaker azimuth location, and was below 0.5 at all test separations below 40° in experiment 1 [Fig. 9(B)].

The reduction of interaural correlation caused by acoustic superposition of lead and lag waveforms would be expected to contribute to the reduction of azimuthal discrimination abilities for lead and lag sources. Saberi *et al.* (1998) measured the effect of decreasing interaural correlation on head turns elicited by dichotic noise bursts in barn owls. The precision of head turns was unaffected as interaural correlation was decreased from 1 to about 0.5. At interaural correlation levels below 0.5 the precision of head turns declined rapidly. Thus, the reductions of interaural correlation caused by acoustic superposition in the present study would be expected to impair spatial resolution. In order to minimize this effect, the owls could have adopted a strategy of listening only to the initial stimulus segment, in the case of lead discrimination, or to the final stimulus segment, in the case of lag discrimination. However, their ability to isolate the ITD information conveyed by the initial and final segments from that conveyed by the overlap segment would be limited by the temporal resolution of the binaural pathway. Wagner (1991) demonstrated that, in a lateralization task, barn owls average ITD information within a ''temporal window'' with equivalent rectangular duration ranging from 3 to 42 ms. If the owls processed directional information with a temporal window of similar duration in the present study, spatial resolution would have been reduced as a result of averaging degraded ITD information conveyed by the overlap segment with the uncorrupted ITD information conveyed by either initial or final segments. This effect is illustrated in Fig. 10, which shows the time-dependent value of binaural correlation computed within two double exponential time windows of the form

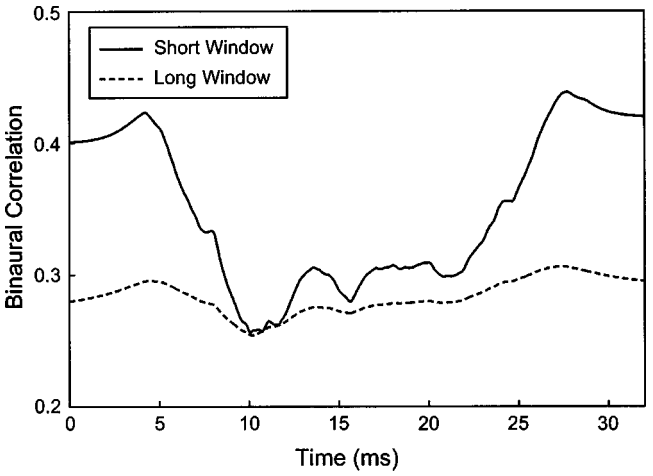


FIG. 10. Binaural correlation values computed within sliding time windows applied to the lead–lag stimulus combination illustrated in Figs. 8 and 9. The shapes of the short (solid line) and long (dotted line) time windows were given by the weighted sum of two exponentials using parameters (short: $T_p=1$ ms, $T_s=12$ ms, $w=0.83$; long: $T_p=3$ ms, $T_s=64$ ms, $w=0.81$) that provided the best fits to behavioral data from the two subjects in the study by Wagner (1991; see the text for equation). Windowing was accomplished by using the time window as a weighting function that was multiplied by the stimulus waveforms prior to calculation of the binaural cross correlation. Binaural correlation values were computed as the time-window center was advanced in 33-μs increments, beginning approximately 2 ms prior to stimulus onset, and ending approximately 2 ms after stimulus offset.

$$W(t) = (1-w)^* \exp(-t/T_p) + w^* \exp(-t/T_s), \quad (2)$$

where T_p and T_s are time constants and w is a weighting factor. The short (solid line; $T_p = 1$ ms, $T_s = 12$ ms, $w = 0.83$) and long (dotted line; $T_p = 3$ ms, $T_s = 64$ ms, $w = 0.81$) time windows used in this illustration were generated using parameters that provided the best fits to the behavioral data for the two owls in Wagner's study, and had equivalent rectangular durations of 3 and 14.5 ms, respectively. In both cases, the binaural correlation values decline as the time window center nears the temporal midpoint of the stimulus, and the maximal levels of binaural correlation occurring near stimulus onset and offset are lower than those measured using rectangular time windows to isolate the initial and final stimulus segments [Fig. 9(A)]. Thus, temporal averaging should limit the ability of owls to make use of the ITD information conveyed by the initial and final stimulus segments. The effect of acoustic superposition on ITD cues illustrated here may contribute to the impairment of discrimination performance for lead and lag sources relative to single sources. It cannot explain the difference in performance between lead-and lag-source conditions, however, because the ITD cues to the locations of both sources are equally affected.

B. Localization dominance

A common interpretation of results obtained in lateralization and discrimination studies of the precedence effect is that lead and lag sounds are combined, perceptually, to form a single, fused event. The perceived lateral position of the fused event reflects a sum of the ITDs of lead and lag sources, weighted so that the lead source makes the dominant contribution. As a result, the amount by which the ITD of one of the sources must be changed to produce a just-detectable change in the perceived location of the fused event is greater for the lag source than it is for the lead source. This interpretation is formalized in a model proposed by Shinn-Cunningham *et al.* (1993), that uses a single metric to summarize lead-source dominance. According to this model, the perceived lateral position, α, is a noise-corrupted weighted sum of the ITDs of the leading (τ_1) and lagging (τ_2) sounds

$$\alpha = c\,\tau_1 + (1-c)\,\tau_2 + \eta, \quad (3)$$

where η is internal noise. The perceptual weighting factor, c, quantifies the extent to which the laterality of the lead sound dominates the laterality of the perceptual image.

This model can be used to calculate the perceptual weighting factor from the discrimination data obtained in the present study with free-field stimulus presentation, following an approach detailed by Litovsky and Macmillan (1994). Because ITD is directly proportional to sound source azimuth, Eq. (2) can be used to derive the following equations that relate the change in perceived azimuthal location of fused images ($\delta\alpha$) to the change in the azimuth of either lead ($\delta \mathrm{Az}_1$) or lag($\delta \mathrm{Az}_2$) sources:

$$\delta\alpha = c\,\delta \mathrm{Az}_1, \quad (4a)$$

$$\delta\alpha = (1-c)\,\delta \mathrm{Az}_2. \quad (4b)$$

TABLE II. Estimates of c from discrimination thresholds measured in experiment 1.

Subject	Single/Lead Eq. (5a)	Single/Lag Eq. (5b)	Lead/Lag Eq. (5c)
889	0.50	0.71	0.67
882	0.67	0.58	0.61
896	0.27	0.83	0.63

The discriminability, d', of a difference in azimuth is defined as the difference in means of the distributions of perceived azimuths for two locations in units of their common standard deviations. The discriminability of a change in azimuth for single, lead, and lag sources is thus

$$d'_{\mathrm{single}} = \delta \mathrm{Az}_{\mathrm{single}}/s, \quad (5a)$$

$$d'_{\mathrm{lead}} = c\,\delta \mathrm{Az}_1/s, \quad (5b)$$

$$d'_{\mathrm{lag}} = (1-c)\,\delta \mathrm{Az}_2/s, \quad (5c)$$

where s is the standard deviation of α. We assume that the variance of perceptual judgments is equal in all three conditions. Because the threshold criterion level of discriminability is the same for all test conditions, the right-hand expressions of any pair of Eqs. (5) can be set equal to one another to give three equations that relate the discrimination thresholds obtained in different test conditions ($t_{\mathrm{single}}, t_{\mathrm{lead}}, t_{\mathrm{lag}}$) in terms of c

$$t_{\mathrm{single}}/t_{\mathrm{lead}} = c, \quad (6a)$$

$$t_{\mathrm{single}}/t_{\mathrm{lag}} = 1-c, \quad (6b)$$

$$t_{\mathrm{lead}}/t_{\mathrm{lag}} = (1-c)/c. \quad (6c)$$

Estimates of c calculated using Eqs. (6) and the threshold measurements from experiment 1 are listed in Table II. All but two of the nine estimates of c are greater than 0.5, consistent with the view that, overall, the lead source tended to dominate perceived laterality of a fused image. There was considerable variation in the values of c computed for each owl with different equations, and across subjects for Eqs. (6a) and (6b). The values of c computed by Eq. (6c) exhibited the least variation across subjects. For owls 889 and 896 the c values given by Eq. (6a) were lower than those given by Eq. (6c), and the values given by Eq. (6b) were higher than those given by Eq. (6c). This result is expected if some factor, in addition to the pulling effect of the stationary source, acts to reduce discrimination abilities in the two source conditions. A likely cause of the observed discrepancies is the reduction of binaural correlation resulting from acoustic superposition. For this reason, we believe that the c values given by Eq. (6c), which does not require comparison of thresholds in one- and two-source conditions, provides the best estimate of perceptual weighting in the present discrimination task.

Shinn-Cunningham *et al.* (1993) demonstrated that the value of c depends on several stimulus parameters, including sound duration and lag delay. The values of c calculated from our data for barn owls are lower, on average, than those obtained in studies using short duration stimuli, including 6-ms noise bursts (Litovsky and Macmillan, 1994), and

clicks (Wallach *et al.*, 1949; Gaskell, 1983; Yost and Soderquist, 1984; Saberi and Perrott, 1990; Shinn-Cunningham *et al.*, 1993). The barn owl *c* values are comparable to those calculated by Shinn-Cunningham *et al.* (1993) from data obtained by Zurek (1980) in a pointer matching experiment, using continuous noises as stimuli at lag delays from 3 to 9 ms. Thus, it is possible that the low *c* values obtained in the present study reflect similar effects of stimulus duration to those in humans.

Several mechanisms have been proposed to account for lead-source dominance in localization. Some authors have proposed that neuronal inhibitory mechanisms triggered by a leading sound suppress the neuronal representation of directional information conveyed by later-arriving sounds (Harris *et al.*, 1963; Lindemann, 1986; Zurek, 1987). Neurophysiological data, showing that addition of a leading sound reduces the responses of binaural or space-tuned neurons to their preferred stimuli, have been interpreted as evidence for such inhibitory processes (Yin, 1994; Fitzpatrick *et al.*, 1995; Keller and Takahashi, 1996; Litovsky and Yin, 1998; Fitzpatrick *et al.*, 1999). Recent modeling studies, however, suggest potential alternative explanations. Tollin (1998) demonstrated that lead-source dominance could result from a temporal weighting function that emphasizes the most recent monaural inputs at the site of binaural comparison, in the absence of any onset-triggered inhibitory mechanism. In biological terms, this weighting function could reflect the physiological properties of neurons that initially detect ITDs in the mammalian medial superior olive (Goldberg and Brown, 1969; Yin and Chan, 1990) or the avian nucleus laminaris (Carr and Konishi, 1990). More recent modeling studies demonstrated that interactions between lead and lag sounds within peripheral filters, in combination with adaptation of hair cell responses, can account for several classic precedence phenomena reported in psychophysical studies (Hartung and Trahiotis, 2001) as well as the physiological results described above (Trahiotis and Hartung, 2002), without requiring any inhibitory processes. At lag delays shorter than the reciprocal of a filter's bandwidth, the lag sound will arrive while the filter is still responding to the lead sound. Interactions between responses to the lead and lag sound result in new ''internal, effective ITDs'' that are dominated by the ITD of the lead sound. The behavioral results obtained in the present study are compatible with any of these proposed mechanisms with the likely exception of interactions within peripheral filters.

It is unlikely that peripheral interactions account for the impaired performance in the lead- and lag-source conditions of the present experiment, because response times of filters in the frequency range of the stimuli are too short for substantial interactions to occur. The passband of stimuli used in this experiment (3–9 kHz) was chosen to encompass the 4 to 8-kHz range over which sound localization by barn owls is most accurate (Konishi, 1973; Knudsen and Konishi, 1979). To assess the effects of peripheral filtering in this frequency range, we used a bank of gamma-tone filters (Slaney, 1993), similar to those used by Hartung and Trahiotis (2001), but with center frequencies ranging from 2670 to 9514 Hz in 1/12-octave increments and proportional bandwidth adjusted to better approximate that of barn owl auditory-nerve fibers (Köppl, 1997). The impulse response of the 2670-Hz filter reaches maximum amplitude at 1.27 ms and decays by 12 dB at a response time of 3 ms. In the range of frequencies from 4 to 8 kHz, the attenuation relative to peak of impulse response amplitudes measured at 3 ms ranged from 20 to 41 dB. Thus, unlike filters tuned to the low frequencies that humans use for ITD extraction, the ringing response of filters tuned to frequencies that barn owls use for sound localization is substantially damped at the 3-ms lag delay used in this experiment.

Nevertheless, to illustrate the effect of peripheral filtering on the ''internal ITD'' expected in the present experimental conditions, a running interaural cross correlogram was computed after first passing the left- and right-ear signals through the filter bank, computing the interaural cross correlogram within each frequency channel, and then averaging the results across frequencies. The resulting, ''filtered'' cross correlogram is shown in Fig. 8(C). Most of the major features of the unfiltered cross correlogram are preserved, despite the temporal smearing effects of the filter responses. In particular, during the overlap segment, intermittently appearing peaks corresponding to lead and lag sources are clearly evident. Thus, the damped ringing of peripheral filters at a delay of 3 ms is of insufficient amplitude to obscure the ITD cues to either lead- or lag-source locations. Note that this analysis likely overestimates the effects of peripheral filtering on azimuthal resolution, because the lower frequency bands with the longest response times are weighted equally to the higher frequencies, which make the greatest contribution to sound localization in owls. The relative insensitivity to changes in lag-source azimuth is therefore more likely the result of either a central inhibitory mechanism that partially suppresses the neuronal representation of lag-source location, asymmetric temporal weighting at the site of binaural interaction, adaptation of hair cell responses, or some combination of these mechanisms. Sufficient information is not currently available to evaluate the contribution of the latter three mechanisms to behavioral precedence phenomena in barn owls.

C. Discrimination of nonspatial cues

Results of experiment 2 demonstrate that barn owls are not able to detect the monaural cues associated with changes of lag-source location, including changes in rippling of the combined signal spectrum. Sensitivity to rippled noise stimuli has been assessed in several nonhuman species. In the echolocating bat, *Eptesicus fuscus*, sensitivity to changes in spectral ripples is sufficient to account for detection of changes in distance between two simulated target surfaces (Simmons *et al.*, 1990). This ability presumably reflects specializations of the bat's auditory system for target ranging and acoustical imaging along the range axis. Thresholds for detection of spectral ripples against a flat-spectrum reference condition have been measured in chinchillas (Shofner and Yost, 1995) and budgerigars (Amagai *et al.*, 1999). In budgerigars, the only avian species to be studied, thresholds for detection of spectral ripples are similar, or slightly lower than those of humans for repetition pitches in the range from

100 to 1000 Hz, corresponding to lead–lag delays of 1 to 10 ms (Amagai *et al.*, 1999). A direct comparison to results of the present study is not possible, however, as sensitivity to changes in lag delay, or repetition pitch, was not measured.

Konishi and Kenuk (1975) demonstrated that barn owls are able to detect smaller differences in signal spectra than the differences between spectra of habituating and test stimuli used in experiment 2. We can suggest two possible explanations for the difference in discrimination performance in the two studies. First, the owl's ability to detect differences between spectral ripples may be restricted to low ripple frequencies. The procedure that Konishi and Kenuk (1975) used to generate spectral differences would have generated an effective spectral ripple with a period of 2/3 octave, which is much longer than the period of ripples tested in the present study. Sensitivity to spectral ripples has been shown to decrease as ripple period is decreased in other birds (Amagai *et al.*, 1999) as well as in humans (Summers and Leek, 1994). Second, long integration times may be required for owls to discriminate fine differences in spectra or temporal structure. The 25- and 100-ms stimuli used in the present study were shorter than those used by Konishi and Kenuk (1975) by approximately 3 orders of magnitude.

D. Implications of discrimination suppression for physiological studies

A major goal of this study was to provide quantitative behavioral measures of the precedence effect for comparison with neuronal response data. Previous single-unit neurophysiological studies have described potential neuronal correlates of the suppression of directional information conveyed by lag sources in precedence effect paradigms (Yin, 1994; Fitzpatrick *et al.*, 1995; Keller and Takahashi, 1996; Litovsky *et al.*, 1997; Litovsky and Yin, 1998; Fitzpatrick *et al.*, 1999). It has long been appreciated, however, that although reflections are not necessarily heard as separate events they contribute to the perceived quality and location of the fused event (Blauert, 1997). More recently, it was demonstrated that humans are impaired, but still able to detect changes in the location of lagging sounds presented in the free field (Perrott *et al.*, 1989; Litovsky and Macmillan, 1994). Results of the present study demonstrate that barn owls are similar to humans in this regard. These findings suggest that, although the neuronal representation of the direction of a lagging source might be largely attenuated by the presence of a leading source, some residual representation must remain in order to provide a basis for detecting changes in its location. Results of a recent study of single-unit responses to binaural click pairs in several auditory structures of the rabbit are consistent with this proposal (Fitzpatrick *et al.*, 1999). Although responses of the majority of neurons to a lagging click at best ITD were suppressed by the presence of a lead click, a small proportion of neurons in each structure responded to the lag click at lead–lag delays from 1 to 4 ms. The findings of the present study demonstrate that barn owls exhibit similar lag- and lead-discrimination suppression effects to humans, and provide a basis for quantitative comparison between behavioral and neuronal discrimination of sound source location. Future experiments will make use of the topographic auditory spatial representation in the barn owl's auditory midbrain to define the neuronal representation of sound sources in simulated echoic conditions, and to relate neuronal measures of discrimination performance to the present behavioral data.

ACKNOWLEDGMENTS

We thank Dr. Kip Keller for providing valuable assistance in writing the computer code used for stimulus generation and data acquisition and Dr. K. Hartung for providing many insightful comments, including a firm definition of the term ''suppression.'' Supported by NIDCD grants DC00448 to M.W.S. and DC03925 to T.T.T.

Amagai, S., Dooling, R. J., Shamma, S., Kidd, T. L., and Lohr, B. (**1999**). ''Detection of modulation in spectral envelopes and linear-rippled noises by budgerigars (*Melopsittacus undulatus*),'' J. Acoust. Soc. Am. **105**, 2029–2035.

Bala, A. D., and Takahashi, T. T. (**2000**). ''Pupillary dilation response as an indicator of auditory discrimination in the barn owl,'' J. Comp. Physiol. [A] **186**, 425–434.

Blauert, J. (**1997**). *Spatial Hearing* (MIT Press, Cambridge, MA).

Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (**1992**). ''The analysis of visual motion: A comparison of neuronal and psychophysical performance,'' J. Neurosci. **12**, 4745–4765.

Carr, C. E., and Konishi, M. (**1990**). ''A circuit for detection of interaural time differences in the brain stem of the barn owl,'' J. Neurosci. **10**, 3227–3246.

Cranford, J. L. (**1982**). ''Localization of paired sound sources in cats: Effects of variable arrival times,'' J. Acoust. Soc. Am. **72**, 1309–1311.

Fitzpatrick, D. C., Kuwada, S., Batra, R., and Trahiotis, C. (**1995**). ''Neural responses to simple simulated echoes in the auditory brain stem of the unanesthetized rabbit,'' J. Neurophysiol. **74**, 2469–2486.

Fitzpatrick, D. C., Kuwada, S., Kim, D. O., Parham, K., and Batra, R. (**1999**). ''Responses of neurons to click-pairs as simulated echoes: Auditory nerve to auditory cortex,'' J. Acoust. Soc. Am. **106**, 3460–3472.

Gaskell, H. (**1983**). ''The precedence effect,'' Hear. Res. **12**, 277–303.

Goldberg, J. M., and Brown, P. B. (**1969**). ''Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization,'' J. Neurophysiol. **32**, 613–636.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Harris, G. G., Flanagan, J. L., and Watson, B. J. (**1963**). ''Binaural interaction of a click with a click pair,'' J. Acoust. Soc. Am. **35**, 672–678.

Hartung, K., and Trahiotis, C. (**2001**). ''Peripheral auditory processing and investigations of the precedence effect which utilize successive transient stimuli,'' J. Acoust. Soc. Am. **110**, 1505–1513.

Keller, C. H., Hartung, K., and Takahashi, T. T. (**1998**). ''Head-related transfer functions of the barn owl: Measurement and neural responses,'' Hear. Res. **118**, 13–34.

Keller, C. H., and Takahashi, T. T. (**1996**). ''Responses to simulated echoes by neurons in the barn owl's auditory space map,'' J. Comp. Physiol. [A] **178**, 499–512.

Kelly, J. B. (**1974**). ''Localization of paired sound sources in the rat: Small time differences,'' J. Acoust. Soc. Am. **55**, 1277–1284.

Knudsen, E. I., and Konishi, M. (**1979**). ''Mechanisms of sound localization by the barn owl (*Tyto alba*),'' J. Comp. Physiol. [A] **133**, 13–21.

Konishi, M. (**1973**). ''Locatable and nonlocatable acoustic signals for barn owls,'' Am. Nat. **107**, 775–785.

Konishi, M., and Kenuk, A. S. (**1975**). ''Discrimination of noise spectra by memory in the barn owl,'' J. Comp. Physiol. **97**, 55–58.

Köppl, C. (**1997**). ''Frequency tuning and spontaneous activity in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, *Tyto alba*,'' J. Neurophysiol. **77**, 364–377.

Lindemann, W. (**1986**). ''Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wave front,'' J. Acoust. Soc. Am. **80**, 1623–1630.

Litovsky, R. Y. (**1997**). ''Developmental changes in the precedence effect: Estimates of minimum audible angle,'' J. Acoust. Soc. Am. **102**, 1739–1745.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). ''The precedence effect,'' J. Acoust. Soc. Am. **106**, 1633–1654.
Litovsky, R. Y., and Macmillan, N. A. (**1994**). ''Sound localization precision under conditions of the precedence effect: Effects of azimuth and standard stimuli,'' J. Acoust. Soc. Am. **96**, 752–758.
Litovsky, R. Y., Rakerd, B., Yin, T. C., and Hartmann, W. M. (**1997**). ''Psychophysical and physiological evidence for a precedence effect in the median sagittal plane,'' J. Neurophysiol. **77**, 2223–2226.
Litovsky, R. Y., and Shinn-Cunningham, B. G. (**2001**). ''Investigation of the relationship among three common measures of precedence: Fusion, localization dominance, and discrimination suppression,'' J. Acoust. Soc. Am. **109**, 346–358.
Litovsky, R. Y., and Yin, T. C. (**1998**). ''Physiological studies of the precedence effect in the inferior colliculus of the cat. I. Correlates of psychophysics,'' J. Neurophysiol. **80**, 1285–1301.
Middlebrooks, J. C., Makous, J. C., and Green, D. M. (**1989**). ''Directional sensitivity of sound-pressure levels in the human ear canal,'' J. Acoust. Soc. Am. **86**, 89–108.
Moiseff, A. (**1989**). ''Bi-coordinate sound localization by the barn owl,'' J. Comp. Physiol. [A] **164**, 637–644.
Musicant, A. D., Chan, J. C., and Hind, J. E. (**1990**). ''Direction-dependent spectral properties of cat external ear: New data and cross-species comparisons,'' J. Acoust. Soc. Am. **87**, 757–781.
Perrott, D. R., Marlborough, K., Merrill, P., and Strybel, T. Z. (**1989**). ''Minimum audible angle thresholds obtained under conditions in which the precedence effect is assumed to operate,'' J. Acoust. Soc. Am. **85**, 282–288.
Poganiatz, I., Nalken, I., and Wagner, H. (**2001**). ''Sound-localization experiments with barn owls in virtual space: influence of interaural time difference on head-turning behavior,'' J. Assoc. Res. Otolaryngol. **2**, 1–21.
Saberi, K., and Perrott, D. R. (**1990**). ''Lateralization thresholds obtained under conditions in which the precedence effect is assumed to operate,'' J. Acoust. Soc. Am. **87**, 1732–1737.
Saberi, K., Takahashi, Y., Konishi, M., Albeck, Y., Arthur, B. J., and Farahbod, H. (**1998**). ''Effects of interaural decorrelation on neural and behavioral detection of spatial cues,'' Neuron **21**, 789–798.
Shinn-Cunningham, B. G., Zurek, P. M., and Durlach, N. I. (**1993**). ''Adjustment and discrimination measurements of the precedence effect,'' J. Acoust. Soc. Am. **93**, 2923–2932.
Shofner, W. P., and Yost, W. A. (**1995**). ''Discrimination of rippled-spectrum noise from flat-spectrum wideband noise by chinchillas,'' Aud. Neurosci. **1**, 127–138.
Simmons, J. A., Moss, C. F., and Ferragamo, M. (**1990**). ''Convergence of temporal and spectral information into acoustic images of complex sonar targets perceived by the echolocating bat, *Eptesicus fuscus*,'' J. Comp. Physiol. [A] **166**, 449–470.
Slaney, M. (**1993**). ''An efficient implementation of the Patterson-Holdsworth auditory filter bank,'' Apple Computer Technical Report 35.
Spezio, M. L., Keller, C. H., Marrocco, R. T., and Takahashi, T. T. (**2000**). ''Head-related transfer functions of the Rhesus monkey,'' Hear. Res. **144**, 73–88.
Summers, V., and Leek, M. R. (**1994**). ''The internal representation of spectral contrast in hearing-impaired listeners,'' J. Acoust. Soc. Am. **95**, 3518–3528.
Tollin, D. J. (**1998**). ''Computational model of the lateralization of clicks and their echoes,'' in NATO Advanced Study Institute on Computational Hearing, edited by Greenberg, S. and Slaney M., pp. 77–82.
Tollin, D. J., and Henning, G. B. (**1998**). ''Some aspects of the lateralization of echoed sound in man. I. The classical interaural-delay based precedence effect,'' J. Acoust. Soc. Am. **104**, 3030–3038.
Trahiotis, C., and Hartung, K. (**2002**). ''Peripheral auditory processing, the precedence effect and responses of single units in the inferior colliculus,'' Hear. Res. **168**, 55–59.
Wagner, H. (**1991**). ''A temporal window for lateralization of interaural time difference by barn owls,'' J. Comp. Physiol. [A] **169**, 281–289.
Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**). ''The precedence effect in sound localization,'' Am. J. Psychol. **57**, 315–336.
Wyttenbach, R. A., and Hoy, R. R. (**1993**). ''Demonstration of the precedence effect in an insect,'' J. Acoust. Soc. Am. **94**, 777–784.
Yin, T. C. (**1994**). ''Physiological correlates of the precedence effect and summing localization in the inferior colliculus of the cat,'' J. Neurosci. **14**, 5170–5186.
Yin, T. C., and Chan, J. C. (**1990**). ''Interaural time sensitivity in medial superior olive of cat,'' J. Neurophysiol. **64**, 465–488.
Yost, W. A., Hill, R., and Perez-Falcon, T. (**1978**). ''Pitch and pitch discrimination of broadband signals with rippled power spectra,'' J. Acoust. Soc. Am. **63**, 1166–1175.
Yost, W. A., and Soderquist, D. R. (**1984**). ''The precedence effect: Revisited,'' J. Acoust. Soc. Am. **76**, 1377–1383.
Zurek, P. M. (**1980**). ''The precedence effect and its possible role in the avoidance of interaural ambiguities,'' J. Acoust. Soc. Am. **67**, 952–964.
Zurek, P. M. (**1987**). ''The Precedence Effect,'' in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer, New York), pp. 85–105.

Evaluation of array-processing algorithms for a headband hearing aid

Julie E. Greenberg,[a] Joseph G. Desloge,[b] and Patrick M. Zurek[b]
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 11 December 2001; revised 13 September 2002; accepted 18 November 2002)

Several array-processing algorithms were implemented and evaluated with experienced hearing-aid users. The array consisted of four directional microphones mounted broadside on a headband worn on the top of the listener's head. The algorithms included two adaptive array-processing algorithms, one fixed array-processing algorithm, and a reference condition consisting of binaural directional microphones. The algorithms were evaluated under conditions with both one and three independent noise sources. Performance metrics included quantitative speech reception thresholds and qualitative subject preference ratings for ease-of-listening measured using a paired-comparison procedure. On average, the fixed algorithm improved speech reception thresholds by 2 dB, while the adaptive algorithms provided 7–9-dB improvement over the reference condition. Subjects judging ease-of-listening generally preferred all array-processing algorithms over the reference condition. The results suggest that these adaptive algorithms should be evaluated further in more realistic acoustic environments. [DOI: 10.1121/1.1536624]

PACS numbers: 43.66.Ts, 43.60.Lq [MRL]

I. INTRODUCTION

Hearing-aid users commonly complain of difficulty understanding speech in the presence of background noise (Kochkin, 2000a, b). Microphone-array hearing aids provide a promising solution to this problem when the noise, or interference, arises from directions other than that of the desired source. These devices process inputs from multiple microphones to produce a single output with improved signal-to-noise ratio (SNR).

Array-processing algorithms can be classified as either *fixed* or *adaptive*. Fixed algorithms provide a static directional response pattern, while adaptive algorithms dynamically alter their directional response depending on the directional properties of the input signals. Fixed array-processing algorithms generally provide performance that is robust with respect to variations in the acoustic environment, with the advantage of low computational complexity. For a given microphone configuration, they can provide near-optimal performance in extreme reverberation and a moderate level of performance in less reverberant environments. Adaptive array-processing algorithms potentially provide superior performance, particularly when a small number of interferers are present in low-to-moderate reverberation, but this comes at the expense of increased computational complexity. Furthermore, under some conditions adaptive algorithms may degrade the desired signal, leading to reduced intelligibility.

Much effort has been given to the design and evaluation of algorithms for microphone-array hearing aids (Greenberg and Zurek, 1992; Soede *et al.*, 1993a, b; Hoffman *et al.*, 1994; Kompis and Dillier, 1994; DeBrunner and McKinney, 1995; Kates and Weiss, 1996; Liu and Sideman, 1996; Schweitzer *et al.*, 1996; Saunders and Kates, 1997; Wittkop, 1997; Desloge, 1998; Kompis *et al.*, 1998; Shields and Campbell, 1998; Vanden Berghe and Wouters, 1998; Gibian *et al.*, 1999; Hoffman *et al.*, 2000; Widrow, 2000; Zurek and Greenberg, 2000). A comprehensive review of such algorithms is available elsewhere (Greenberg and Zurek, 2001).

When evaluating microphone-array hearing aids, it is important to choose an appropriate reference condition. In order to justify the increase in physical size and/or computational complexity required for such devices, evaluations of more complex systems should be based on comparisons to simpler microphone-array systems. Therefore, the reference condition for evaluating a fixed array-processing algorithm should be either a single directional microphone or binaural directional microphones.[1] Similarly, the reference condition for an adaptive array-processing algorithm should be a comparable fixed system, that is, fixed array processing of the signals obtained from microphones in the same physical configuration. This approach permits quantification of the incremental benefits due to each increase in the physical size and/or computational complexity of the device.

The performance of both fixed and adaptive array-processing algorithms is highly dependent on the acoustic environment, specifically on the position and number of noise sources and the level, latency, and duration of reverberation. It is therefore difficult to make meaningful comparisons across studies, because the acoustic characteristics of the test environments have strong effects on array performance. Even so, results of previous studies can be broadly summarized as follows. Fixed processing of the signals from multiple microphones forming relatively large arrays (10–20 cm) can improve speech reception thresholds (SRTs) by 5–10 dB over those obtained with a single omnidirectional microphone, or by 2–3 dB over a single directional microphone (e.g., Soede, 1993b; Saunders and Kates, 1997; Widrow, 2000). Because of their physical size and visibility,

[a]Electronic mail: jgreenbe@mit.edu
[b]Present address: Sensimetrics Corporation, 48 Grove Street, Somerville, MA 02144.

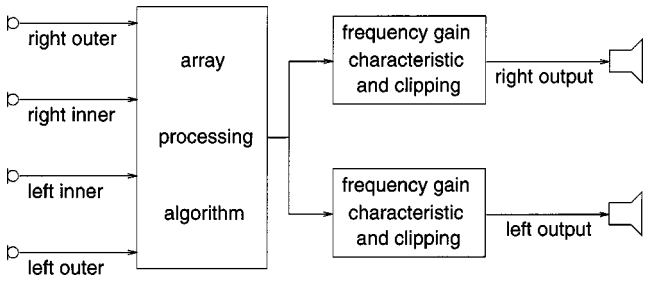


FIG. 1. Block diagram of experimental hearing aid system for evaluating array-processing algorithms. Four microphone signals comprise the inputs to the array-processing algorithms, described in Sec. II. Each algorithm produces a single output signal, which is processed by frequency gain characteristics and clipping levels determined for each ear as described in Sec. III C.

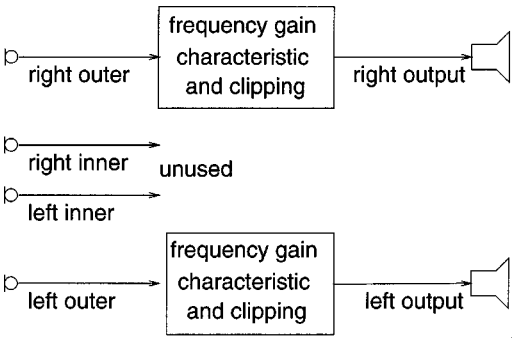


FIG. 2. Block diagram of experimental hearing aid system for the reference condition. The two outermost microphone signals are processed by frequency gain characteristics and clipping levels determined for each ear as described in Sec. III C. The inner microphone signals are not used.

these arrays may be cosmetically unacceptable to a portion of the hearing-impaired population. A more discreet option is provided by adaptive processing of the signals from two microphones worn at the ears. With one independent noise source in mild reverberation, such systems can improve SRTs by 10 dB over single or binaural directional microphones (Kompis *et al.*, 1998; Zurek and Greenberg, 2000) and by more than 5 dB over fixed processing of signals from the same physical array (Kompis and Dillier, 1994).

Acoustic theory and simulation results predict that adaptive arrays can provide substantial benefits only in relatively nonreverberant environments. In strong reverberation, the benefits of microphone arrays are due to their simple directionality, and basic acoustic principles dictate that the directionality of an array increases with its size. Therefore, an adaptive algorithm in conjunction with a relatively large microphone array can potentially provide the best performance for the hearing-aid user under the widest range of acoustic conditions.

The current study explores a microphone-array configuration intended to expand the range of performance versus cosmetic trade-offs that might be offered to hearing-aid users. It consists of four microphones mounted on a headband worn atop the head. Directional microphones are used to achieve rejection of rearward sounds and to obtain the complementary benefits in terms of directionality generally provided by directional microphones as elements of larger arrays (Stadler and Rabinowitz, 1993; Kompis and Dillier, 1994; Kompis *et al.*, 1998).

In this study, three array-processing algorithms (two adaptive and one fixed) were compared to a reference condition that used binaural directional microphones. Previous work has already addressed the development, refinement, and preliminary evaluation of these algorithms with physical measurements and listening by normal-hearing subjects (Greenberg and Zurek, 1992; Greenberg, 1994; Welker *et al.*, 1997; Desloge *et al.*, 1997; Desloge, 1998). The current goal is to move beyond initial evaluation of the algorithms themselves and to assess the benefit that they might provide to hearing-impaired individuals. Accordingly, this study employed experienced hearing-aid users as subjects, with careful attention to fitting of the frequency responses to ensure audibility of target speech materials. The evaluation included both speech reception thresholds in noise and relative ease of listening. The algorithms were tested in a mildly reverberant environment in order to explore any potential advantage of the adaptive algorithms, since adaptive algorithms are not expected to provide much benefit over comparable fixed systems in strong reverberation. Noise configurations included roving noise sources and multiple directional interferers (independent noise sources arriving from different spatial directions). A real-time system permitted the subjects to wear the microphone array and listen to the processed outputs while sitting in the sound field.

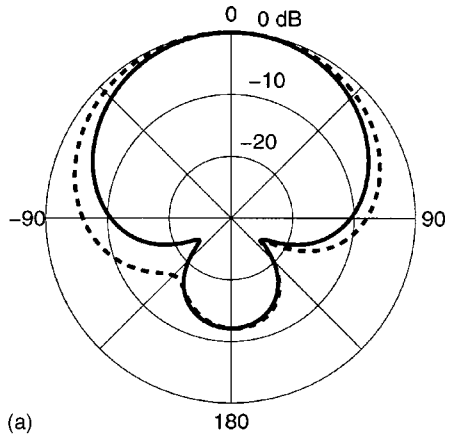
II. HEARING AID ALGORITHMS

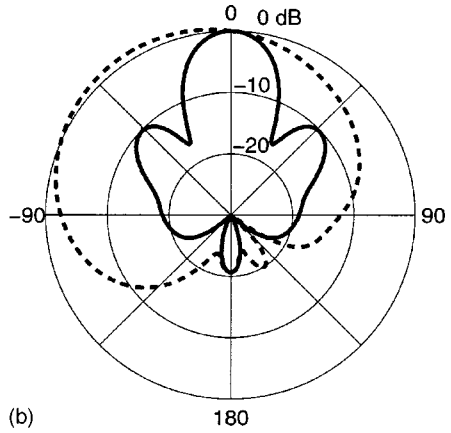
All algorithms used an array of four microphones mounted on a headband (described in Sec. III A) and were designed on the assumption that the desired signal arrives from array broadside. All algorithms except the reference condition produce a single output, which is then processed for presentation to both ears (Fig. 1).

(i) REF—For the reference condition, signals from the outermost microphones form left and right output signals for binaural presentation (Fig. 2). Binaural directional microphones were selected for the reference condition because they are cosmetically acceptable, commercially available, and provide the most directionality compared to all other options with wide commercial availability. Figure 3 (dashed lines) shows the theoretical directional response of the left channel of the reference condition.

(ii) SUM4—For this fixed-processing algorithm, the four microphone signals are simply averaged together (Fig. 4). This algorithm represents delay-and-sum weighting for the broadside case; it can be implemented without a special-purpose processor. Figure 3 (solid lines) shows the theoretical directional response of the SUM4 algorithm.

(iii) GJI—This adaptive algorithm is based on the modified Griffiths–Jim beamformer (Greenberg and Zurek, 1992). The primary signal consists of the average of the four microphone signals (which, by itself, is identical to the SUM4 output), while the reference signal, which is subsequently processed adaptively, is the difference between the two innermost[2] microphone sig-





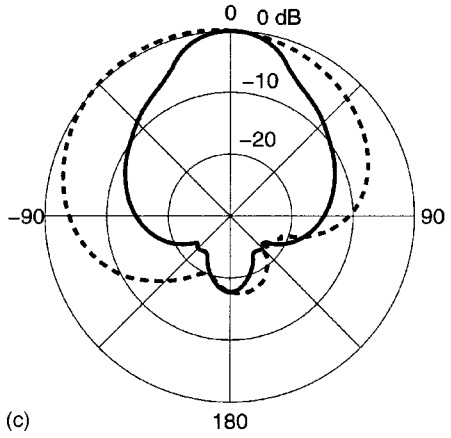


FIG. 3. Polar responses for the left channel of the REF condition (dashed) and the SUM4 fixed array processor (solid). The plots show the theoretical magnitude response (normalized to the 0-degree direction) for the headband array placed on a rigid sphere with 9-cm radius. (a) Responses at 500 Hz, corresponding to directivity indices of 5.2 dB (REF) and 5.9 dB (SUM4). (b) Responses at 3000 Hz, corresponding to directivity indices of 4.2 dB (REF) and 10.3 dB (SUM4). (c) Average intelligibility-weighted (Greenberg *et al.*, 1993) responses over the range 200–4000 Hz, corresponding to intelligibility-weighted directivity indices of 4.6 dB (REF) and 8.1 dB (SUM4).

nals (Fig. 5). Because this algorithm employs one adaptive filter based on two microphone signals, it is capable of steering one independent null and is expected to be most effective in environments with a single directional interferer. Further detail is given in Sec. II A.

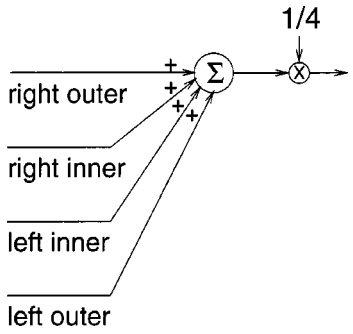


FIG. 4. Block diagram of SUM4 array-processing algorithm. The four microphone signals are averaged.

(iv) LENS—This adaptive algorithm is a narrow-band technique that estimates optimal null locations to minimize interferer power and then steers nulls accordingly. Because all four microphone signals are processed adaptively, this algorithm can steer three independent nulls and is expected to be effective in environments with up to three directional interferers. Further detail is given in Sec. II B.

A. Adaptive algorithms based on the modified Griffiths–Jim beamformer (GJI)

Figure 5 shows a block diagram of the modified Griffiths–Jim beamformer implemented in this study. This system includes two modifications to the design that was originally proposed by Griffiths and Jim (1982), the sum method of LMS weight update and correlation-based inhibition, both described below. These modifications are designed to mitigate problems encountered in the presence of strong desired signals (Greenberg and Zurek, 1992).

The primary signal (upper channel) is formed from the average of the four microphone signals, and then delayed. The reference signal is formed from the difference between the two innermost microphone signals. The reference signal is processed by an adaptive filter; the adaptive filter weights are updated using the LMS algorithm (Widrow and Stearns, 1985). The filtered reference signal is then subtracted from the delayed primary signal to form the system output.

1. Sum method of LMS weight update

The filter weights of the adaptive beamformer are continuously updated in order to track variations in the acoustic

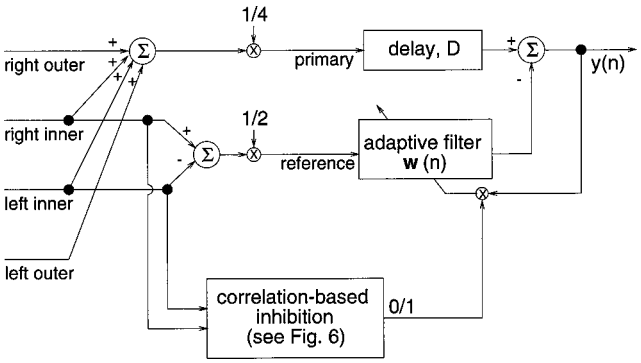


FIG. 5. Block diagram of GJI array-processing algorithm. This adaptive algorithm is described in Sec. II A.

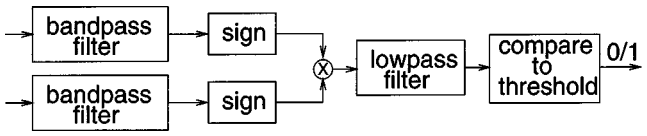


FIG. 6. Block diagram of correlation-based inhibition described in Sec. II A 2.

environment. The LMS algorithm is an attractive choice for adjusting the adaptive weights because it is easy to implement and computationally simple. The sum method of LMS weight update[3] (Greenberg, 1998) is given by

$$\mathbf{w}(n+1)=\mathbf{w}(n)+\mu(n)y(n)\mathbf{x}(n), \tag{1}$$

with

$$\mu(n)=\frac{\alpha}{L[\hat{\sigma}_y^2(n)+\hat{\sigma}_x^2(n)]}, \tag{2}$$

where n is the time index, $\mathbf{w}(n)$ is the $L\times 1$ vector of adaptive filter weights, L is the adaptive filter length, $y(n)$ is the system output (quantity to be minimized), $\mathbf{x}(n)$ is the data vector containing the L samples in the adaptive filter's tapped delay line at time n, and α is a dimensionless step-size parameter. The power estimates $\hat{\sigma}_x^2(n)$ and $\hat{\sigma}_y^2(n)$ are time-varying estimates of the adaptive filter input and system output signal powers, respectively. These estimates were obtained by squaring the instantaneous signal values and then smoothing with a first-order recursive low-pass filter. The time constant of the low-pass filter used to compute $\hat{\sigma}_x^2(n)$ is selected to correspond to the length of the adaptive filter, while the time constant used to compute $\hat{\sigma}_y^2(n)$ is set to 10 ms in order to track power fluctuations of speech. The current study used values of $\alpha=0.01$ and $L=285$, with a 16-kHz sampling frequency.

2. Correlation-based inhibition

The correlation-based inhibition (Greenberg and Zurek, 1992) utilizes running measures of the zero-lag correlation coefficients between the innermost pair of microphones. The goal is to identify intervals when the desired speech signal is relatively strong and then refrain from updating the adaptive weights during those intervals. This prevents cancellation of the desired signal and reduces fluctuations in the filter weights due to the desired signal.

The correlation-based inhibition operates as shown in Fig. 6. The two innermost microphone signals are bandpass filtered. The filter's center frequency is selected to minimize the expected correlation from directional sources arising within the range of nondesired angles for that microphone spacing (Greenberg, 1994). The bandpass filter used in this study was a 45-point finite impulse response filter designed using the Parks–McClellan method; its passband was between 1.7 and 3.4 kHz (selected for the 6-cm microphone spacing) with transition bands of 500 Hz.

The bandpass filtered signals are hard limited to ± 1, and the instantaneous correlation for each sampling interval is computed from the product of two hard-limited values. The instantaneous correlation is then smoothed by a first-order recursive low-pass filter with a time constant of 10 ms; this time constant is much longer than the sampling period required for speech signals, but short enough to track power fluctuations corresponding to the duration of individual phonemes. Finally, the low-pass-filtered correlation value is compared to the threshold value of zero, which has been found to perform well in a variety of acoustic environments (Greenberg, 1994). If the correlation value is below the threshold, then the estimated SNR is low and the adaptive weights are updated for the current iteration according to (1). If the correlation value is above the threshold, the estimated SNR is high, and the weights are not updated for that iteration, that is,

$$\mathbf{w}(n+1)=\mathbf{w}(n). \tag{3}$$

B. The location-estimating, null-steering (LENS) algorithm

The LENS algorithm differs from the modified Griffiths–Jim beamformer in several ways:

(1) It applies adaptive filters directly to the microphone input signals, rather than to a set of reference signals.
(2) It is a narrow-band algorithm; that is, the inputs are block-processed using the FFT, and LENS processing is applied independently at each FFT bin.
(3) It inserts an intermediate step in the determination of the adaptive filters. Specifically, it uses the array input signals to determine a set of LENS parameters, which may be interpreted as interferer source location estimates when the microphones are arranged in a uniform linear array. After robustness-control processing to limit potential degradation of the desired signal, these parameters are used to formulate the actual adaptive filters.

This section provides a concise description of the LENS algorithm for an arbitrary number of microphones (M). Figure 7 shows a block diagram of the LENS algorithm with $M=4$. A more complete description and mathematical derivation is available elsewhere (Desloge, 1998).

The main premise of the LENS algorithm is that it is possible to parametrize the array processor's adaptive filters in a way that describes useful information about the array processing. For each processing block, the M adaptive filters at the discrete frequencies k may be represented as an M-element weight vector $\underline{W}(k)=[W_0(k)W_1(k)\cdots W_{M-1}(k)]^T$. LENS is based upon a parametrization of $\underline{W}(k)$ developed in Desloge (1998) for uniform linear arrays. Specifically, LENS constructs the weights, $\underline{W}(k)$, as a function of $M-1$ complex-valued LENS parameters $\underline{\beta}(k)=[\beta_0(k)\beta_1(k)\cdots\beta_{M-2}(k)]^T$. The LENS parametrization is designed so that the following properties hold:

(i) $\sum_{i=0}^{M-1}W_i[\underline{\beta}(k)]=1$ regardless of the values of the parameters $\beta_j(k)$. This means that the weight vector, $\underline{W}[\underline{\beta}(k)]$, always preserves sources arriving from array broadside, that is, sources that appear equally at all microphones.
(ii) Each parameter, $\beta_j(k)$, controls one independent null in the free-field directional response formed by apply-

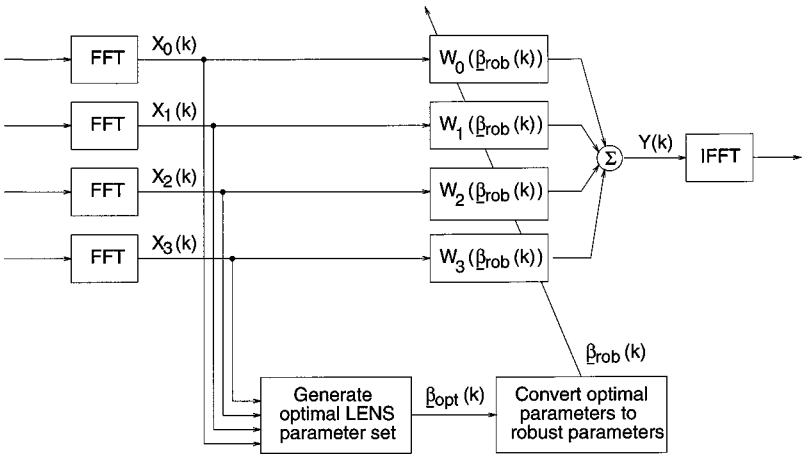


FIG. 7. Block diagram of LENS array-processing algorithm. This adaptive algorithm is described in Sec. II B.

ing $\underline{W}[\underline{\beta}(k)]$ to a uniform linear array. Therefore, by modifying a single parameter, $\beta_j(k)$, it is possible to shift a single null in the response pattern without affecting the locations of the remaining nulls controlled by the other LENS parameters, $\beta_i(k)$, where $i \neq j$.

In general, an M-element weight vector can steer $M-1$ independent nulls while preserving one desired signal direction. For the LENS algorithm, the constraint to preserve the desired signal is hard-wired into the functional dependence of the weight vector, $\underline{W}[\underline{\beta}(k)]$, on the parameters, $\underline{\beta}(k)$. The $M-1$ LENS parameters each control one of the $M-1$ independent nulls.

Given this parametrization, the narrow-band array-processing problem [that of choosing $\underline{W}(k)$ to minimize expected output power subject to the constraint of preserving sources from a specific location] can be reformulated as a problem of choosing the "optimal" $\underline{\beta}(k)$ so that $\underline{W}[\underline{\beta}(k)]$ minimizes the expected output power. Suppressing the explicit k dependence for simplicity, this may be expressed mathematically by defining $\underline{X}=[X_0 X_1 \cdots X_{M-1}]^T$ to be the set of narrow-band array inputs and expressing the output filtering and summing operations as the inner product $Y[\underline{\beta}] = \underline{W}[\underline{\beta}]^H \underline{X}$. In this case,

$$\underline{\beta}_{\text{opt}} = \underset{\underline{\beta} \in \mathbf{C}^{M-1}}{\text{argmin}}\, E[|Y[\underline{\beta}]|^2] = \underset{\underline{\beta} \in \mathbf{C}^{M-1}}{\text{argmin}}\, \underline{W}[\underline{\beta}]^H R_{XX} \underline{W}[\underline{\beta}], \tag{4}$$

where $E[\ ]$ denotes expected value, $R_{XX}=E[\underline{X}\underline{X}^H]$ is the input covariance matrix, and $\mathbf{C}^{M-1}$ is the $M-1$-dimensional vector space of complex numbers. Since $\underline{W}[\underline{\beta}]$ is formulated to preserve the desired (broadside) signal direction for any $\underline{\beta} \in \mathbf{C}^{M-1}$, the resulting $\underline{W}[\underline{\beta}_{\text{opt}}]$ will steer nulls towards interference sources.

Once $\underline{\beta}_{\text{opt}}$ has been determined, it is then evaluated and modified in order to identify and prevent potential nulling and degradation of the desired signal. This is necessary because misalignment or reverberation violates the assumption of the desired signal arriving from array broadside, potentially allowing the LENS optimization of (4) to steer a null towards the desired source. However, since each LENS parameter, $\beta_{\text{opt},j}$, controls one independent null steered by the system, it is possible to observe $\underline{\beta}_{\text{opt}}$ and to identify any nonrobust parameters that are steering nulls towards the misaligned or reverberated desired source. These nonrobust parameters are then set to zero,[4] while leaving the remaining parameters unchanged, in order to shift potentially degrading nulls away from the desired source direction without affecting the other nulls. The resulting robust parameter set, $\underline{\beta}_{\text{rob}}$, leads to the final adaptive filters used by the LENS algorithm, given by $\underline{W}[\underline{\beta}_{\text{rob}}]$.

For the implementation used in this study, the inputs were FFT processed using 256-sample blocks that overlapped by 128 samples. The sequence of FFT blocks was used to form running estimates of the input covariance matrices R_{XX} at each frequency. These R_{XX} estimates were then used in an iterative, relaxation-based technique to determine $\underline{\beta}_{\text{opt}}$, the optimal narrow-band LENS parameter minimization of (4). This result was then evaluated to identify any parameters that steered nulls within 30 degrees of the desired signal direction at array broadside. The robust parameter set, $\underline{\beta}_{\text{rob}}$, was formed by setting the corresponding elements of $\underline{\beta}_{\text{opt}}$ to zero, leading to the final adaptive filter weights $\underline{W}[\underline{\beta}_{\text{rob}}]$.

III. EXPERIMENTAL PROCEDURES

A. Hearing aid system

The hearing aid system used in this study consisted of microphones, preamplifier circuits, digital signal processing boards installed in a personal computer, and hearing aid receivers. Subjects wore a headband holding four forward-pointing cardioid microphones (Knowles EL-3085) in broadside orientation with 6-cm spacing. The microphone signals were preamplified by 36 dB and then presented to the analog inputs of two DSP-96 boards (Ariel Corporation). Each of these boards contains 16-bit stereo A/D and D/A and a Motorola 96002 processor. The boards were programmed to sample the microphone signals at 16 kHz, perform the array processing algorithms described in Sec. II, and apply frequency-dependent amplification and digital clipping appropriate for each ear. (The fitting procedure is described in Sec. III C.) The analog output signals were presented to the subject via Knowles receivers (either ED-1932 or CI-2748, depending on the subject's hearing loss) embedded in custom full shell in-the-ear (ITE) modules with 1.5-mm-diam vents. A vent plug was used to reduce the vent to 0.5 mm for most

subjects. However, for two subjects (HK and PG), the desired gains could only be provided without feedback when a closed plug was used to completely occlude the vent.

B. Test environment

The experiments were performed in a double-wall soundproof room with internal dimensions of $2.7\times2.5\times2.0$ m. The walls and ceiling were perforated metal panels with acoustically absorptive foam backing. The floor was a solid metal panel covered with a carpet. The space is relatively nonreverberant, but not anechoic; acoustic measurements indicated a lower bound of 5 dB for the direct-to-reverberant ratio at 1.0 m. The reverberation time (T60) of the room measured in octave bands was 200 ms at 125 Hz, 100 ms at 250 Hz, and 70–80 ms at higher frequencies. The subject, wearing the microphones and receivers, sat in the center of the booth. Cables connected the microphones and receivers to a preamplifier. Test stimuli were delivered from an array of seven loudspeakers (Radio Shack Optimus Pro 7) placed at a height of 1.1 m and a distance of 1.0 m from the center of the booth in the directions 0, ±45, ±90, ±135 degrees. The experimenter, the computer containing the DSP boards, and the test equipment were located outside the booth. The subject used a hand-held terminal and an intercom to communicate with the experimenter and to record responses.

C. Hearing aid fitting

The experimental hearing aid system was programmed to provide amplification appropriate for each ear and to prevent the presentation of uncomfortably loud sounds. Frequency-gain characteristics were implemented by 127-point finite impulse response filters determined via the procedure described below. The outputs of the experimental hearing aid system were limited by simple digital clipping of the output signals whenever the computed output values exceeded thresholds based on measurements of the subject's uncomfortable loudness levels (Cox, 1995).

The fitting process used in this study has been described previously (Greenberg *et al.*, 2000). The goal of this fitting process is to determine binaural frequency-gain characteristics based on a compromise between matching the insertion gains prescribed by NAL-R (Byrne and Dillon, 1986) and matching the insertion gains provided by the subject's personal hearing aids. The frequency-gain characteristics are based on

(i) the NAL-R target insertion gain, computed from measurements of the subject's unmasked thresholds,
(ii) real-ear insertion gain of the subject's personal hearing aids with the volume control set by the subject for comfortable listening of conversational speech in quiet,
(iii) real-ear insertion gain of the experimental hearing aid system with known flat digital gain as the frequency response, and
(iv) feedback path transfer functions computed from programming the hearing aid to play a known random noise sequence and then measuring the received digital signal in the absence of any acoustic input.

These quantities are all functions of frequency and are used to create two initial binaural frequency-gain characteristics. The first one matches the insertion gains of the subject's personal hearing aids (PHA) and the second matches the maximum at each frequency of PHA and the insertion gains prescribed by NAL-R. The rationale for this choice was that subjects were likely to accept frequency-gain characteristics that matched their own aids, and might accept additional gain beyond that provided by their personal aids when prescribed by NAL-R. It was assumed that subjects would not accept reductions in gain relative to their personal aids, even if those reductions were prescribed by NAL-R. All of the frequency-gain characteristics are subject to the constraint that the gain at each frequency is limited to 5 dB below the level resulting in instability as predicted by the feedback path transfer functions.

During the fitting procedure, the subject listens to conversational speech from a straight-ahead source using the two outermost cardioid microphones, which correspond to the binaural reference condition. The subject first listens via the PHA responses in order to verify that it "sounds like" their own aids. Next, the subject listens via the maximum PHA/NAL responses and is asked if that would be acceptable for listening for extended periods. If the subject accepts, then the maximum PHA/NAL response is used for the study. If not, a new frequency-gain characteristic halfway between PHA and maximum PHA/NAL is created and the subject is asked again. This process continues until the subject accepts a set of binaural frequency-gain characteristics, or until the difference between successive frequency gain characteristics is less than 5 dB for both ears and all frequencies. The result of this process is that the subject receives the strongest acceptable binaural frequency-gain characteristics with maximum gains at each frequency limited by the insertion gains of either their personal hearing aids or the NAL-R target. Although the directional filtering provided by the algorithms may alter the frequency response for off-axis interferers, all of the algorithms preserve this fitted frequency response for desired signals arriving from straight-ahead.

D. Noise conditions

For all noise conditions, the desired speech signal was delivered from the straight-ahead (0 degrees) loudspeaker. The interfering sources were delivered from one or more of the off-axis loudspeakers, depending on the noise condition:

(i) NS1—One interferer (speech-shaped noise) was fixed at $+45$ degrees.
(ii) NS2—One interferer (speech-shaped noise) roved among the six off-axis locations (±45, ±90, and ±135 degrees). The interferer switched to a new location after a random time interval uniformly distributed between 1.5 and 4.5 s.
(iii) NS3—Three fixed interferers consisted of speech-shaped noise at -45 degrees, a female talker at $+90$ degrees (Mayes, 1998), and a male talker at $+135$

degrees (Rosaforte, 1997). The relative levels of these three sources were adjusted to be equal as measured at the center of the speaker array.

(iv) NS4—Two fixed interferers consisted of a male talker at +45 degrees (Rosaforte, 1997) and a female talker at +90 degrees (Mayes, 1998). An additional interferer (speech-shaped noise) roved among the remaining four off-axis locations (−45, −90, and ±135 degrees). The roving source switched to a new location after a random time interval uniformly distributed between 1.5 and 4.5 s. The relative levels of these three sources were adjusted to be equal as measured at the center of the speaker array.

E. Test procedures

The algorithms were tested for speech intelligibility under the two noise conditions with fixed sources (NS1 and NS3) and for ease of listening under all four noise conditions.

1. Speech intelligibility

Speech reception thresholds (SRTs) were measured using Harvard sentences (IEEE, 1969) read by a male talker as the desired signal. The sentences were presented at a fixed level (either 55, 60, or 65 dB SPL) determined to insure audibility for each subject. The interference levels were varied as a part of the adaptive procedure for measuring SRTs.

Each SRT measurement was based on a list of seven sentences, with the first sentence repeated at successively higher signal-to-noise ratios (SNRs) adjusted in 4-dB increments, until the subject correctly repeated at least three of the five keywords. For presentation of successive sentences, the SNR was decremented by 2 dB if the subject correctly repeated at least three of the five keywords in the previous sentence. The SNR was incremented by 2 dB if the subject failed to repeat at least three of the keywords in the previous sentence. The estimate of SRT for a single run was determined from the average SNR used to present sentences 2–8 of each list. (Although an eighth sentence was never presented, the SNR for the eighth sentence is calculated from the subject's response to the seventh sentence.)

The interference sources were activated 1 s before the onset of each sentence, so the adaptive algorithms could at least partially converge. After each sentence presentation, the sources were turned off so the subject could speak his/her response via intercom for the experimenter to record. For each condition (algorithm and noise condition) the SRT was measured twice.

2. Ease of listening

Relative ease of listening was assessed with a paired-comparison procedure. The desired signal was continuous speech material read by a male talker (Packard, 1995) and presented at 65 dB SPL. The interferers were in one of the four noise configurations described above. For NS1 and NS2, the single interferer was presented at 70 dB SPL, for a SNR of −5 dB. For NS3 and NS4, the levels of all three interferers were adjusted so that their combined level was 65 dB SPL, for an SNR of 0 dB.

TABLE I. Description of subjects.

Subject	Gender	Age	Receiver	Speech level for SRT (dB SPL)
WF	M	78	ED	55
MG	M	75	ED	60
PG	M	59	CI	60
GI	M	64	ED	55
HK	F	72	CI	65
TM	M	73	ED	55
GP	F	54	ED	65
AS	M	75	ED	55
RS	M	71	ED	55

Each trial compared two of five algorithms,[5] randomly assigned to be "A" and "B." The subject listened through both "A" and "B," switching between them at will, while the source signals played continuously. The subject was given the following instructions and asked to select a rating corresponding to the scale shown.

> You will be comparing two hearing aids, A and B. When the trial begins, you will be presented with speech from speaker 4 and one or more interfering sounds from speakers 1, 2, 3, 5, 6, or 7. Switch into algorithm A or B by pressing the A or B key. Try to understand the speech using algorithms A and B. Switch between algorithms as many times as you like. Decide if A or B makes the speech easier to listen to, or you may decide that listening is equally easy with A or B. Use keys 1–5 to indicate your preference between A and B.
>
> 1—A much better than B
> 2—A slightly better than B
> 3—No noticeable difference
> 4—B slightly better than A
> 5—B much better than A.

A run consisted of 20 trials (2 repetitions×10 unique pairs of the 5 algorithms). The noise condition was fixed for each run, and each run was repeated on a different day, for a total of 8 runs (2 repetitions×4 noise conditions). The order of trials was randomized within a run, and the order of runs was also randomized.

F. Subjects

Table I describes the nine subjects who participated in this study, and Fig. 8 shows their audiograms. All subjects had bilaterally symmetric sensorineural hearing loss and passed tympanometry screening, except for one subject (GI) whose right ear had a mild conductive component below 1000 Hz and showed reduced compliance. All subjects had at least three years experience wearing hearing aids binaurally.

IV. RESULTS

A. Speech intelligibility

Figure 9(a) shows the measured SRTs, averaged over subjects and repetitions, for each noise condition individually and averaged over noise conditions. Since the SRT re-

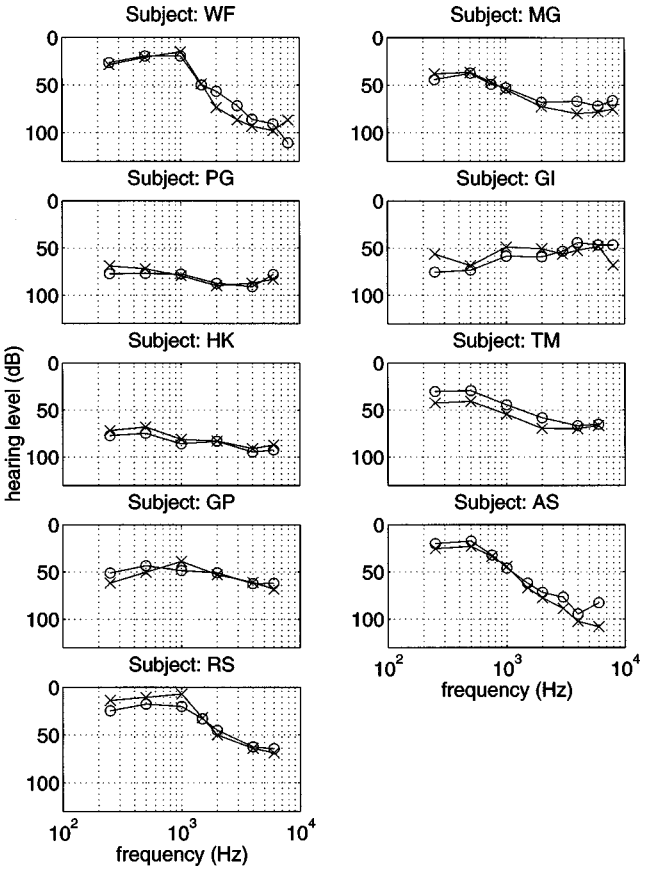


FIG. 8. Audiograms of the nine subjects participating in this study.

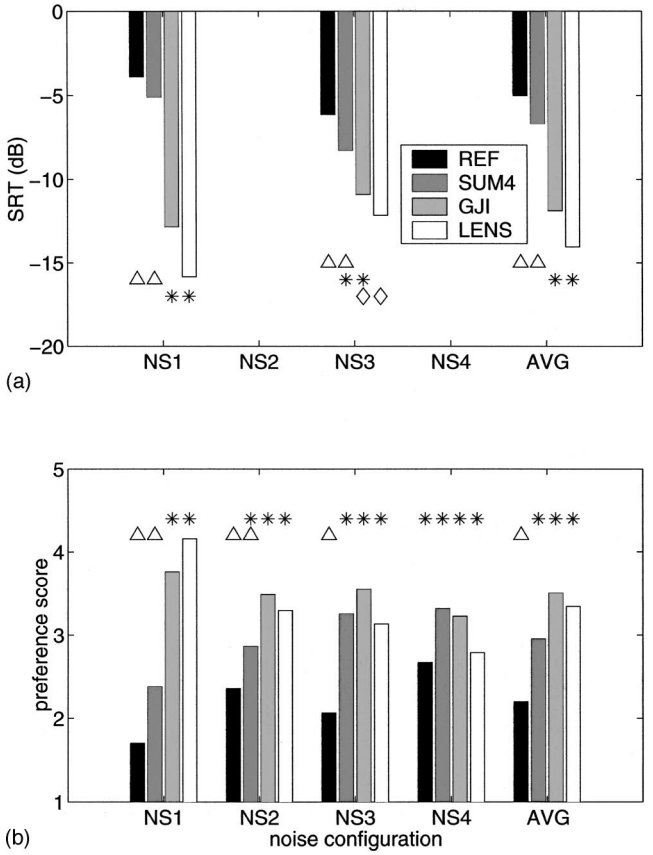


FIG. 9. Results of listening tests, averaged across subjects and repetitions, for noise conditions described in Sec. III D. For each set of bars, algorithms with the same symbols above/below the bars were not significantly different according to a *post hoc* Tukey HSD test. (a) Average speech reception thresholds (SRTs) for each algorithm. (b) Average preference scores for each algorithm as determined from the ease-of-listening ratings.

flects the SNR at which the sentence material was 50% intelligible, lower SRT values indicate better performance. Subject-specific data for each noise condition are presented in Table II, which shows the change in SRTs for each algorithm relative to the reference condition. These data were averaged over repetitions. Negative values indicate better performance of the array processing algorithms over the reference condition.

On the average, array processing improved SRTs by 2, 7, and 9 dB over the reference condition, for the SUM4, GJI, and LENS algorithms, respectively. The adaptive algorithms (GJI and LENS) generally outperformed both the reference condition (REF) and the fixed array-processing algorithm (SUM4). The fixed array processing algorithm performs slightly better than the reference condition for most subjects. Within the adaptive algorithms, LENS performs somewhat better than GJI for most subjects and noise conditions. Although the magnitudes of the improvements vary with subject, no obvious subject-specific trends are evident.

A repeated-measures analysis of variance (ANOVA) on the SRT data analyzed three fixed factors: algorithm, noise condition, and repetition number. Significant ($P<0.0001$) main effects of algorithm were found, as well as significant ($P=0.0002$) algorithm×noise configuration interaction. No other main effects or interactions were significant. A *posthoc* Tukey HSD test showed significant differences ($P=0.05$) between two algorithm groupings: (REF and SUM4) and (GJI and LENS), as indicated by the symbols below the rightmost set of bars in Fig. 9(a).

The algorithm×noise configuration interaction was found to be a relatively minor effect due to the fact that the adaptive algorithms (GJI and LENS) produced lower SRTs for NS1 than NS3, while the other two algorithms (REF and SUM4) produced lower SRTs for NS3 than NS1. The differences in mean SRTs between noise conditions for the same algorithm were 2–4 dB. Repeating the *posthoc* Tukey HSD test for each noise configuration individually produced the same algorithm groupings for NS1 as for the averaged data, but slightly different groupings for NS3, as shown in Fig. 9(a).

B. Ease of listening

Ease-of-listening results for each noise condition, averaged over subjects and repetitions, are presented in Tables III–VI. Table VII presents the same data averaged over the four noise conditions. The ratings were adjusted to reflect presentation as algorithm B, regardless of the actual order of presentation.[6] Higher scores indicate preference for algorithm B over algorithm A on a greater number of trials, according to the rating scale described in Sec. III E 2.

Average preference scores for each algorithm are presented in Fig. 9(b). These preference scores were determined by averaging the ratings for an algorithm (presented as algorithm B) when compared to the other three algorithms (pre-

TABLE II. Average change in SRTs in dB due to each algorithm, relative to the reference condition, for each subject and noise configuration. Negative values indicate that the array-processing algorithm improved performance over the reference condition.

Subject	Algorithm	NS1	NS3
HK	SUM4	2.0	−2.0
	GJI	−2.0	−5.3
	LENS	−5.4	−7.6
GI	SUM4	−0.0	−2.6
	GJI	−20.9	−9.2
	LENS	−19.4	−14.1
PG	SUM4	0.1	−1.1
	GJI	−7.1	−2.9
	LENS	−14.9	−7.1
TM	SUM4	−6.6	−2.0
	GJI	−16.0	0.0
	LENS	−20.3	−2.9
WF	SUM4	−4.6	2.3
	GJI	−9.7	−1.1
	LENS	−9.1	−1.4
AS	SUM4	0.3	−1.7
	GJI	−4.6	−2.6
	LENS	−14.3	−4.3
GP	SUM4	−2.6	−6.0
	GJI	−5.3	−9.1
	LENS	−5.3	−6.6
MG	SUM4	2.6	−3.1
	GJI	−5.4	−5.3
	LENS	−7.4	−7.3
RS	SUM4	−2.0	−2.9
	GJI	−9.4	−7.3
	LENS	−11.1	−3.9

sented as algorithm A). This is equivalent to averaging the values in each column of Tables III–VII. Note that because of the comparative nature of the preference scores, the absolute scores are only meaningful for relating algorithms within a particular noise configuration; there is no direct interpretation to the observation that, for example, the preference score of the reference algorithm increases from NS1 to NS2.

Subject-specific data for each noise condition are presented in Table VIII, which shows the change in preference score for each algorithm relative to the reference condition. These data were averaged over repetitions. Positive values indicate preference for the array-processing algorithms over the reference condition.

TABLE III. Ease-of-listening results for noise condition NS1, averaged over subjects and repetitions. Each numerical entry indicates the average rating for pairwise comparisons of algorithms A and B.

		ALG B			
		REF	SUM4	GJI	LENS
A	REF		3.53	4.64	4.72
L	SUM4	2.47		3.86	4.53
G	GJI	1.36	2.14		3.22
A	LENS	1.28	1.47	2.78	

TABLE IV. Ease-of-listening results as in Table III, for noise condition NS2.

		ALG B			
		REF	SUM4	GJI	LENS
A	REF		3.36	3.98	3.59
L	SUM4	2.64		3.46	3.31
G	GJI	2.02	2.54		2.98
A	LENS	2.41	2.69	3.02	

The ease-of-listening data indicate that all subjects (with the possible exception of MG) generally preferred any array-processing algorithm over the reference condition for all noise configurations. For noise configurations with only one interference source (NS1 and NS2), most subjects preferred the adaptive algorithms (GJI and LENS) over the fixed array-processing algorithm (SUM4). Results are mixed for noise configurations with multiple interferers (NS3 and NS4), and none of the array-processing algorithms is preferable to the others.

The preference scores (averaged over repetitions but not subjects) were submitted to a repeated-measures ANOVA with two fixed factors: algorithm and noise condition. This analysis revealed significant ($P<0.0001$) main effects of algorithm and algorithm×noise configuration interaction. The main effect of noise configuration was not significant. A *posthoc* Tukey HSD test showed significant differences ($P=0.05$) between two algorithm groupings: (REF) and (SUM4, GJI, and LENS), as indicated by the symbols above the rightmost set of bars in Fig. 9(b).

The algorithm×noise configuration interaction can be examined in Fig. 9(b), along with the results of repeating the *posthoc* Tukey HSD test for each noise configuration individually. These results can be understood by considering how each algorithm operates in the presence of different noise configurations, as discussed in the following section.

V. DISCUSSION

A. Noise configurations

With one stationary interferer at +45 degrees and low SNRs, noise configuration NS1 presents a situation where the adaptive algorithms are expected to provide substantial interference reduction. This expectation is confirmed by the experimental results. For the ease-of-listening experiment with a fixed SNR of −5 dB, both adaptive algorithms are substantially preferred over the reference condition and the

TABLE V. Ease-of-listening results as in Table III, for noise condition NS3.

		ALG B			
		REF	SUM4	GJI	LENS
A	REF		3.78	4.21	3.81
L	SUM4	2.22		3.06	2.95
G	GJI	1.79	2.94		2.63
A	LENS	2.19	3.05	3.37	

TABLE VI. Ease-of-listening results as in Table III, for noise condition NS4.

		ALG B			
		REF	SUM4	GJI	LENS
A	REF		3.45	3.45	3.09
L	SUM4	2.55		2.99	2.51
G	GJI	2.55	3.01		2.76
A	LENS	2.91	3.49	3.24	

TABLE VII. Ease-of-listening results as in Table III, averaged over all four noise conditions.

		ALG B			
		REF	SUM4	GJI	LENS
A	REF		3.53	4.07	3.81
L	SUM4	2.47		3.34	3.32
G	GJI	1.93	2.66		2.90
A	LENS	2.19	2.68	3.10	

fixed algorithm. This general trend is also reflected in the SRT measurements, where even lower SNRs were used with the adaptive algorithms.

With one roving interferer and the same low SNR as NS1, noise configuration NS2 presents a more challenging situation for the adaptive algorithms, since they must continuously reconverge to the new interference locations. Accordingly, the results for NS2 are less conclusive than for NS1, but still show an advantage for all array-processing algorithms over the reference condition. Although not statistically significant, it appears that the adaptive algorithms may provide some benefit over SUM4. This suggests that the adaptive algorithms do provide interference suppression after convergence each time the noise source moves, but that the suppression is not as dramatically beneficial as in NS1, where repeated convergence is not necessary. It should be noted that the roving noise source in this study changed position quite frequently, providing a worst-case scenario for evaluating issues related to convergence.

Noise configuration NS3 consists of three stationary interferers all at the same level. This presents a situation where LENS is expected to have the best performance, since it has enough degrees of freedom to cancel three independent interferers. With one adaptive filter, GJI is expected to only be effective against one directional interference; with multiple interferers, GJI would attenuate the strongest noise source if a level disparity existed, but it does not in this case. Furthermore, since its primary signal is equivalent to the SUM4 algorithm, the performance of the GJI algorithm is expected to approach that of SUM4 as the number of interferers increases.

These expectations are confirmed by the SRT measurements, but not by the ease-of-listening results. Under NS3, the ease-of-listening experiment only shows that all three array-processing algorithms are significantly better than the reference condition. This result is likely due to a combination of factors. First, the relatively greater preference for SUM4 is explained by the observation that its fixed directional pattern is close to the optimal response for a diffuse sound field. The presence of additional directional interferers decreases the advantage of the adaptive algorithms over the fixed algorithm. Second, although LENS may be providing more interference suppression than GJI, at this moderate SNR of 0 dB, the output is also more prone to artifacts that arise from the two-stage robustness-control processing. If subjects find such artifacts unpleasant, that could explain the relatively good performance of GJI, which provides less interference suppression but also has fewer artifacts. Because the SRTs were generally measured at lower SNRs, it is likely that these artifacts were not as prominent in that experiment.

With two stationary interferers and one roving interferer, NS4 presents the most challenging noise configuration for the adaptive algorithms. Indeed, in this case the ease-of-listening experiment shows no significant difference between any of the algorithms, including the reference condition.

B. Relation to previous results

The results of this study are consistent with results of previous work and extend our understanding of the potential benefits of adaptive array processors for hearing aids. As described in Sec. I, two-microphone adaptive systems in mild reverberation with one interferer improve SRTs by 10

TABLE VIII. Average change in preference score due to each algorithm, relative to the reference condition, for each subject and noise configuration. Positive values indicate that the experimental algorithm received a higher score than the reference condition.

Subject	Algorithm	NS1	NS2	NS3	NS4
HK	SUM4	0.4	0.7	1.5	−0.1
	GJI	1.5	0.9	2.3	0.2
	LENS	2.8	1.8	1.2	0.2
GI	SUM4	0.3	−0.7	0.8	0.9
	GJI	1.8	1.2	1.6	0.8
	LENS	2.7	0.8	1.6	0.9
PG	SUM4	1.1	2.0	2.2	1.6
	GJI	2.8	2.0	3.0	1.7
	LENS	3.5	2.3	1.5	0.1
TM	SUM4	1.0	1.1	1.6	1.7
	GJI	3.3	2.0	1.3	1.7
	LENS	3.1	2.6	1.8	1.7
WF	SUM4	0.5	0.7	0.8	0.1
	GJI	0.6	0.7	1.2	−0.3
	LENS	1.9	0.0	0.4	−0.1
AS	SUM4	1.2	1.6	3.0	−0.0
	GJI	3.0	2.1	2.7	0.5
	LENS	2.5	0.0	2.3	−2.5
GP	SUM4	0.1	−0.1	0.1	1.2
	GJI	2.4	0.7	0.6	−0.6
	LENS	2.5	0.4	0.7	0.4
MG	SUM4	1.2	−0.8	−0.1	−0.1
	GJI	1.3	−0.5	−0.4	0.3
	LENS	0.8	−0.8	−0.5	0.1
RS	SUM4	0.4	0.1	0.9	0.6
	GJI	1.9	1.2	1.1	0.7
	LENS	2.3	1.3	0.7	0.3

dB over binaural directional microphones (Kompis *et al.*, 1998; Zurek and Greenberg, 2000) and by more than 5 dB over a comparable fixed system (Kompis and Dillier, 1994). This is similar to the results for the NS1 configuration in the current study, which showed an average improvement in SRTs of 9–12 dB for the adaptive algorithms relative to binaural directional microphones (REF) and of 8–11 dB relative to comparable fixed processing (SUM4). It is noteworthy that for this noise configuration, the current four-microphone adaptive algorithms did not outperform those two-microphone adaptive systems studied previously. One may conclude that the potential benefits of additional microphones will only be observed in acoustic environments with more than one interferer and/or stronger reverberation.

Kates and Weiss (1996) evaluated a five-microphone array in two moderately reverberant rooms with single and multiple interferers. The adaptive algorithms considered in that study performed similarly to fixed algorithms implemented with the same physical array, making it difficult to justify the increased computational complexity. This is *not* the finding in the current study, which showed a clear benefit for both adaptive algorithms relative to the fixed algorithm. Other researchers have considered a different array-processing algorithm utilizing three to seven microphones. Evaluations in mildly-to-moderately reverberant environments with single and multiple interferers show improvements of roughly 5 dB over a comparable fixed system (Hoffman *et al.*, 1994, 2000). This is similar to the results of the current study showing improvements of 8–11 dB with a single interferer and 3–4 dB with multiple interferers for the adaptive algorithms relative to a comparable fixed processor.

C. Array-processing algorithms

In general, the results of both the speech intelligibility and ease-of-listening experiments show the following in the mildly reverberant acoustic environment used in this study:

(i) All three array-processing algorithms always perform at least as well as, and often better than, the binaural reference condition.
(ii) The fixed array-processing algorithm may provide some benefits over the binaural reference condition, particularly when multiple interferers are present.
(iii) Both adaptive array-processing algorithms provide substantial benefits over the fixed array-processing algorithm and the reference condition.

Differences between the two adaptive array-processing algorithms were not statistically significant for any of the data collected in this study. However, in terms of objective speech intelligibility measures, the LENS algorithm appears slightly superior to GJI, while the ease-of-listening results show a slight preference for GJI over LENS. This relatively weak finding can be attributed to one or both of the following explanations. First, the LENS algorithm may introduce artifacts which do not interfere with intelligibility, but which listeners nevertheless judge to be detrimental to speech quality. Second, this finding may be due to the different SNRs used in the two experiments, since objectionable artifacts caused by the LENS algorithm are more prominent at higher SNRs. Serious consideration should be given to the ease-of-listening results, because the 0 dB SNR used in that experiment reflects everyday listening situations more closely than the extremely low SNRs used in the SRT measurements with the adaptive algorithms.

The results of this study clearly show that adaptive array-processing algorithms can provide substantial benefits for hearing-impaired persons listening to speech in background noise, compared to both a binaural reference condition and a comparable fixed array-processing system. The benefits provided by the four-microphone adaptive systems are similar to those observed with more cosmetically acceptable two-microphone adaptive systems (Kompis *et al.*, 1998; Zurek and Greenberg, 2000), although the two-microphone systems have not been tested in the presence of multiple interferers and/or strong reverberation, conditions where the additional microphones provide a theoretical advantage. Future work should assess the benefits of the headband array configuration with the GJI and LENS adaptive algorithms in more realistic (and more reverberant) acoustic environments and noise conditions encountered by hearing aid users in daily life.

ACKNOWLEDGMENTS

This work was supported by Contract No. N01-DC-5-2017 and by Grant No. NIH-2-R01-DC00117 from the National Institute of Deafness and Other Communicative Disorders. The authors are grateful to A. R. Brughera, W. M. Rabinowitz, M. Brantley, and L. A. Delhorne for their assistance with this project.

[1]Relative to a single omnidirectional microphone, a single directional microphone may improve SRTs by 1–8 dB (Leeuw and Dreschler, 1991; Valente *et al.*, 1995) depending on the acoustic environment. Values of 3–5 dB are typical in moderate reverberation (Hawkins and Yacullo, 1984; Valente *et al.*, 2000).

[2]The *I* in *GJI* refers to the use of the two *innermost* microphones.

[3]This is in contrast to the traditional method of LMS weight update, which does not include the output signal power in the denominator of (2). By including the output signal power in the normalization of the adaptive step size, the sum method prevents large fluctuations in the filter weights when the desired signal is strong.

[4]A value of zero is used because the mathematical definition of β_j is such that parameters that steer nulls near the desired signal direction have relatively large magnitudes, while parameters that steer nulls away from the desired signal direction have smaller magnitudes. A range of nulls symmetrically restricted away from the desired source direction corresponds to a region centered geometrically about $\beta_j=0$ in the complex plane. As a result, $\beta_j=0$ is the most robust value possible for a LENS parameter, in that it is furthest from the non-robust region. The actual physical direction of the null steered by $\beta_j=0$ varies with frequency.

[5]In addition to the four algorithms described in Sec. II, a fifth algorithm was originally included in the experimental procedure. It was subsequently discovered that that algorithm was not operating properly, so those data are not reported. It should be noted that the ease-of-listening experiment compared all possible pairs of the five algorithms (ten pairs), but the results reported in Sec. IV are based only on comparisons among the four algorithms that were operating properly (six pairs).

[6]As a consequence of this adjustment, each table contains six independent values; values opposite each other along the main diagonal sum to six, the total of the maximum and minimum ratings.

Byrne, D. J., and Dillon, H. A. (**1986**). ''The National Acoustic Laboratories (NAL) new procedure for selecting the gain and frequency response of a hearing aid,'' Ear Hear. **7**, 257–265.
Cox, R. M. (**1995**). ''Using loudness data for hearing aid selection: The IHAFF approach,'' Hearing J. **48**, 10–44.
DeBrunner, V. E., and McKinney, E. D. (**1995**). ''A directional adaptive least-mean-square acoustic array for hearing aid enhancement,'' J. Acoust. Soc. Am. **98**, 437–444.
Desloge, J. G. (**1998**). ''The location-estimating null-steering (LENS) algorithm for adaptive microphone-array processing,'' Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.
Desloge, J. G., Rabinowitz, W. M., and Zurek, P. M. (**1997**). ''Microphone-array hearing aids with binaural output—Part I: Fixed-processing systems,'' IEEE Trans. Speech Audio Process. **5**, 529–542.
Gibian, G. L., Koroljow, W., LaRow, A., Shaw, S., Nelson, P., and Sherlock, L. (**1999**). ''Clinical trials of a hybrid adaptive beamformer (HAB) for improved speech understanding in noise,'' Collected Papers, 137th meeting of the Acoustical Society of America and the 2nd Convention of the European Acoustics Association, Forum Acusticum, integrating the 25th German Acoustics DAGA Conference, Berlin, 14–19 March.
Greenberg, J. E. (**1994**). ''Improved design of microphone-array hearing aids,'' Ph.D. thesis, Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA.
Greenberg, J. E. (**1998**). ''Modified LMS algorithms for speech processing with an adaptive noise canceller,'' IEEE Trans. Speech Audio Process. **6**, 338–351.
Greenberg, J. E., and Zurek, P. M. (**1992**). ''Evaluation of an adaptive-beamforming method for hearing aids,'' J. Acoust. Soc. Am. **91**, 1662–1676.
Greenberg, J. E., and Zurek, P. M. (**2001**). ''Microphone-array hearing aids,'' in *Microphone Arrays: Techniques and Applications*, edited by M. S. Brandstein and D. B. Ward (Springer, New York).
Greenberg, J. E., Peterson, P. M., and Zurek, P. M. (**1993**). ''Intelligibility-weighted measures of speech-to-interference ratio and speech system performance,'' J. Acoust. Soc. Am. **94**, 3009–3010.
Greenberg, J. E., Zurek, P. M., and Brantley, M. (**2000**). ''Evaluation of feedback-reduction algorithms for hearing aids,'' J. Acoust. Soc. Am. **108**, 2366–2376.
Griffiths, L. J., and Jim, C. W. (**1982**). ''An alternative approach to linearly constrained adaptive beamforming,'' IEEE Trans. Antennas Propag. **30**, 27–34.
Hawkins, D. B., and Yacullo, W. S. (**1984**). ''Signal-to-noise ratio advantage of binaural hearing aids and directional microphones under different levels of reverberation,'' J. Speech Hear Disord. **49**, 278–286.
Hoffman, M. W., Pinkelman, C., Lu, X. F., and Li, Z. (**2000**). ''Real-time and off-line comparisons of standard array configurations containing three and four microphones,'' J. Acoust. Soc. Am. **107**, 3560–3563.
Hoffman, M. W., Trine, T. D., Buckley, K. M., and Van Tasell, D. J. (**1994**). ''Robust adaptive microphone array processing for hearing aids: Realistic speech enhancement,'' J. Acoust. Soc. Am. **96**, 759–770.
IEEE (**1969**). *IEEE Recommended Practice for Speech Quality Measurements*, Vol. 297 (IEEE, New York).
Kates, J. M., and Weiss, M. R. (**1996**). ''A comparison of hearing-aid array-processing techniques,'' J. Acoust. Soc. Am. **99**, 3138–3148.
Kochkin, S. (**2000a**). ''MarkeTrak V: Consumer satisfaction revisited,'' Hearing J. **53**, 38–55.
Kochkin, S. (**2000b**). ''MarkeTrak V: ''Why my hearing aids are in the drawer: The consumers' perspective,'' Hearing J. **53**, 34–42.
Kompis, M., and Dillier, N. (**1994**). ''Noise reduction for hearing aids: Combining directional microphones with an adaptive beamformer,'' J. Acoust. Soc. Am. **96**, 1910–1913.
Kompis, M., Feuz, P., Valentini, G., and Pelizzone, M. (**1998**). ''A combined fixed/adaptive beamforming noise-reduction system for hearing aids,'' Proc. 20th Int. Conf. IEEE Eng. Med. Biol. Soc., pp. 3136–3139.
Leeuw, A. R., and Dreschler, W. A. (**1991**). ''Advantages of directional hearing aid microphones related to room acoustics,'' Audiology **30**, 330–344.
Liu, C., and Sideman, S. (**1996**). ''Simulations of fixed microphone arrays for directional hearing aids,'' J. Acoust. Soc. Am. **100**, 848–856.
Mayes, F. (**1998**). *Under the Tuscan Sun* (sound recording) (Bantam Doubleday Dell Audio, New York).
Packard, D. (**1995**). *The HP Way: How Bill Hewlett and I Built our Company* (sound recording) (Harper Audio, New York).
Rosaforte, T. (**1997**). *Tiger Woods: The Makings of a Champion* (sound recording) (Books on Tape, Newport Beach, CA).
Saunders, G., and Kates, J. (**1997**). ''Speech intelligibility enhancement using hearing-aid array processing,'' J. Acoust. Soc. Am. **102**, 1827–1837.
Schweitzer, H., Terry, A. M., and Grim, M. A. (**1996**). ''Three experimental measures of a digital beamforming signal processing algorithm,'' J. Am. Acad. Audiol **7**, 230–239.
Shields, P. W., and Campbell, D. R. (**1998**). ''Multi-microphone noise cancellation for improvement of hearing aid performance,'' Proc. Int. Conf. Acoust. Speech Signal Proc. **ICASSP-98**, 3633–3636.
Soede, W., Berkhout, A. J., and Bilsen, F. A. (**1993a**). ''Development of a directional hearing instrument based on array technology,'' J. Acoust. Soc. Am. **94**, 785–798.
Soede, W., Bilsen, F. A., and Berkhout, A. J. (**1993b**). ''Assessment of a directional microphone array for hearing impaired listeners,'' J. Acoust. Soc. Am. **94**, 799–808.
Stadler, R. W., and Rabinowitz, W. M. (**1993**). ''On the potential of fixed arrays for hearing aids,'' J. Acoust. Soc. Am. **94**, 1332–1342.
Valente, M., Fabry, D. A., and Potts, L. G. (**1995**). ''Recognition of speech in noise with hearing aids using dual microphones,'' J. Am. Acad. Audiol **6**, 203–210.
Valente, M., Schuchman, G., Potts, L. G., and Beck, L. B. (**2000**). ''Performance of dual-microphone in-the-ear hearing aids,'' J. Am. Acad. Audiol **11**, 181–189.
Vanden Berghe, J., and Wouters, J. (**1998**). ''An adaptive noise canceller for hearing aids using two nearby microphones,'' J. Acoust. Soc. Am. **103**, 3621–3626.
Welker, D. P., Greenberg, J. E., Desloge, J. G., and Zurek, P. M. (**1997**). ''Microphone-array hearing aids with binaural output—Part II: A two-microphone adaptive system,'' IEEE Trans. Speech Audio Process. **5**, 543–551.
Widrow, B. (**2000**). ''A microphone array for hearing aids,'' in Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium, Lake Louise, Alberta, Canada, 1–4 October, pp. 7–11.
Widrow, B., and Stearns, S. D. (**1985**). *Adaptive Signal Processing* (Prentice–Hall, Englewood Cliffs, NJ).
Wittkop, T., Albani, S., Hohmann, V., Peissig, J., Woods, W. S., and Kollmeier, B. (**1997**). ''Speech processing for hearing aids: Noise reduction motivated by models of binaural interaction,'' Acustica **83**, 684–699.
Zurek, P. M., and Greenberg, J. E. (**2000**). ''Two-microphone adaptive array hearing aids with monaural and binaural outputs,'' in Proceedings of the Ninth IEEE DSP Workshop, Hunt, TX, 15–18 October.

Unsteady flow through *in-vitro* models of the glottis

G. C. J. Hofmans and G. Groot
Fluid Dynamics Laboratory, Faculty of Applied Physics, Eindhoven University of Technology, Postbus 513, 5600 MB Eindhoven, The Netherlands

M. Ranucci and G. Graziani
University "La Sapienza," Dipartimento Di Meccanica E Aeronautica, Via Eudossiana 18, 00184 Rome, Italy

A. Hirschberg[a)]
Fluid Dynamics Laboratory, Faculty of Applied Physics, Cascade 2.24, Eindhoven University of Technology, Postbus 513, 5600 MB Eindhoven, The Netherlands

(Received 20 August 2001; revised 13 November 2002; accepted 12 December 2002)

The unsteady two-dimensional flow through fixed rigid *in vitro* models of the glottis is studied in some detail to validate a more accurate model based on the prediction of boundary-layer separation. The study is restricted to the flow phenomena occurring within the glottis and does not include effects of vocal-fold movement on the flow. Pressure measurements have been carried out for a transient flow through a rigid scale model of the glottis. The rigid model with a fixed geometry driven by an unsteady pressure is used in order to achieve a high accuracy in the specification of the geometry of the glottis. The experimental study is focused on flow phenomena as they might occur in the glottis, such as the asymmetry of the flow due to the Coanda effect and the transition to turbulent flow. It was found that both effects need a relatively long time to establish themselves and are therefore unlikely to occur during the production of normal voiced speech when the glottis closes completely during part of the oscillation cycle. It is shown that when the flow is still laminar and symmetric the prediction of the boundary-layer model and the measurement of the pressure drop from the throat of the glottis to the exit of the glottis agree within 40%. Results of the boundary-layer model are compared with a two-dimensional vortex-blob method for viscous flow. The difference between the results of the simpiflied boundary-layer model and the experimental results is explained by an additional pressure difference between the separation point and the far field within the jet downstream of the separation point. The influence of the movement of the vocal folds on our conclusions is still unclear.
[DOI: 10.1121/1.1547459]

PACS numbers: 43.70.Bk [AL]

I. INTRODUCTION

Researchers in the field of biomechanics have been using numerical simulations as a useful tool for their studies. Topics that are related to the flow of blood through arteries received a lot of attention in recent years (Rosenfeld, 1995; Luo and Pedley, 1996; Pedrizzetti, 1996) [for an overview of recent work on this subject see Pedley and Luo (1998)]. Also the flow of air through the glottis has been the subject of numerical studies and experimental studies. Interest in this topic is motivated by two research aims: one is the development of prosthetic vocal folds (Lous *et al.*, 1998; De Vries, 2000) and the other is the development of artificial speech models. Recently some attempts at numerically simulating the flow through the glottis including forced vocal-fold movement have been carried out (Alipour and Titze, 1996; Alipour *et al.*, 1996). Also the effect of an asymmetry in the glottal channel on speech production has been investigated by means of steady pressure measurements along the glottal channel supported by numerical simulations (Scherer *et al.*, 2001). Although a complete simulation of the vocal-fold movement and the air flow through the glottis can yield some interesting results, it is not a solution to the problem of artificial (real-time) speech modeling in the near future.

One approach in speech modeling is to model the interaction of the air flow through the glottis and the movement of the vocal folds using simplified models. The problems that are encountered are numerous since the flow through the glottis is a result of the coupling between complex fluid dynamics and complex elastic structure (vocal folds) behavior. Usually both aspects are simplified until such a level is reached that artificial real-time speech production is possible. This leads to oversimplifying both aspects of vocal-fold movement but especially the fluid dynamical description has been reduced to a caricature of the actual flow. An important parameter in these flow models is the point at which the airflow separates from the vocal folds. This parameter determines not only the volume flow through the glottis but also the hydrodynamic forces exerted on the vocal folds.

In most models of the flow through the glottis *ad hoc* assumptions are made about the separation point in the glottis. The most well known model of this kind is the model of Ishizaka and Flanagan (1972) in which flow separation at

[a)]Author to whom correspondence should be addressed. Electronic mail: A.Hirschberg@phys.tue.nl

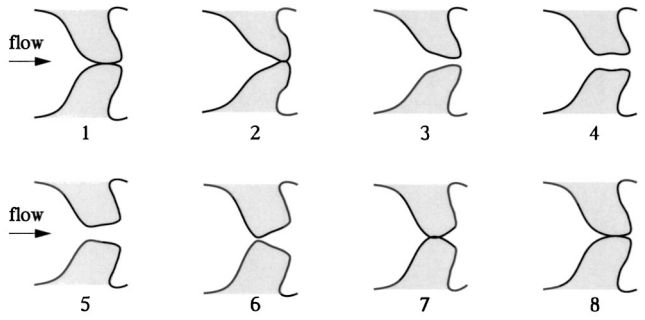


FIG. 1. Typical vocal-fold movement during one oscillation. Note the changing shape of the glottis.

sharp edges is assumed. Such sharp edges are of course not present in the glottis. An attempt was made to improve the description of flow separation within the glottis by using a quasi-steady model based on the boundary-layer theory (Pelorson *et al.*, 1994). It was shown that this improved the results obtained from a simplified mechanical model in which the vocal folds are each represented by two coupled oscillators. This approach is similar to a model proposed for collapsible tubes (Cancelli and Pedley, 1985). The results of this model together with an idea of Liljencrants (1993) formed the basis of a new model (Lous *et al.*, 1998). In the same paper the consequences of the various simplifying assumptions that are made in most models for the vocal-fold movement are also discussed. A discussion on the flow phenomena that might be important for a model of the flow through the glottis can be found in Hirschberg *et al.* (1996), while in Pelorson *et al.* (1997) the focus is on the fluid dynamics of so-called bilabial plosives.

In Fig. 1 a typical cycle of the vocal-fold movement is presented. Note the changing shape of the glottis due to the way in which the vocal folds open and close. Because pressure is applied upstream of the vocal folds, they first start to open at the upstream side (panels 2 and 3 in Fig. 1). This results in a converging glottis shape during the opening phase of the glottis. The closing of the glottis also starts at the upstream side, resulting in a diverging glottis shape during the closing phase (panels 5 and 6 in Fig. 1). Apparently the movement of the vocal-fold tissue at the downstream side is always lagging behind the movement at the upstream side. A glottal pulse that is the result of such a vocal-fold movement representing the sound /a/ at 110 Hz is sketched in Fig. 2 [after data measured by Rothenberg (1981)].

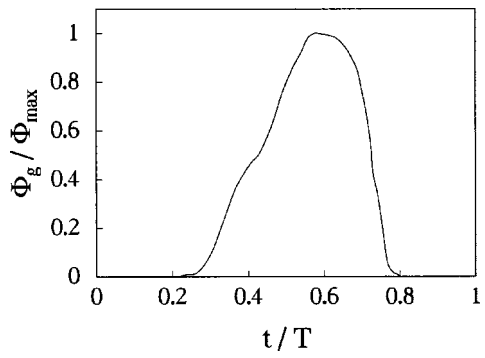


FIG. 2. Normalized glottal flux ϕ_g for the sound /a/ at 110 Hz.

In this paper we present an experimental, theoretical, and numerical study of the flow through *in vitro* models of the glottis. In order to limit ourselves to the fluid dynamical aspects of the flow we used rigid fixed (scale) models of the vocal folds. The size is approximately three times the size of the real vocal folds and the shape of these scale models is inspired by the typical shape of the glottis during the closing phase. In this phase the slowly diverging shape of the glottis implies a rather uncertain position of the separation point. This is the reason why we focus on this geometry. In the actual flow through the glottis an almost steady pressure difference is imposed across the vocal folds, while the opening and closing of the glottis result in an unsteady flow. We decided to use fixed models (not oscillating) for the sake of a high accuracy in the specification of the geometry of the model of the glottis. In order to maintain a similarity to the actual flow through the glottis, an unsteady pressure drop is imposed across the vocal folds. Care is taken that the nondimensional control parameters determining the flow through the glottis have the relevant values. In the glottis these parameters are the Strouhal number and the Reynolds number. The Strouhal number is defined as $\mathrm{Sr}=l/u_0T$ in which l is a stream-wise length scale of the glottis (5–10 mm), u_0 is a typical velocity in the glottis (10–30 m/s), and T is a typical time scale of the vocal-fold movement. For men a typical frequency of oscillation is 10^2 Hz, while for women it is twice as large at approximately 2×10^2 Hz. A better time scale is, however, the opening or closing time of the glottis (2–4 ms). The Reynolds number is defined as $\mathrm{Re}=h_0u_0/\nu$, in which h_0 is the typical height of the glottis (1–2 mm) and ν is kinematic viscosity of air ($\nu=1.5\times10^{-5}$ m^2/s at atmospheric pressure and room temperature). The Strouhal number (typically of order 10^{-1}) is a measure of the influence of unsteady effects on the flow, while the Reynolds number (typically of order 10^3) reflects the importance of viscous effects on the flow. Due to the geometrical scaling factor 3, when the Reynolds and Strouhal analogies are respected, the measured velocities should be multiplied by 3, the pressure should be multiplied by 9, and the time intervals should be divided by 9 in order to compare to physiological conditions. In our experiments Re is in the range of 2×10^3–8×10^3 corresponding to very loud speech. The Reynolds number is also an indication for the onset of turbulence. In the free jet at the exit of the glottis a Reynolds number above 2×10^3 implies turbulence. The turbulence in the glottis upstream of the separation point is however not expected because $\mathrm{Re}\cdot l/h_0 \leqslant 10^5 [l/h_0=\mathcal{O}(10)]$. The Strouhal number in our experiments is of the order of magnitude of 10^{-1}.

In speech modeling the important quantity that has to be modeled accurately is the acoustic source that is the result of the vocal fold movement. This source is the unsteady volume flux through the glottis. In particular the time dependence of this flux during the closing phase of the vocal-fold movement is important because it contains most of the higher harmonics for which our hearing is most sensitive. This movement is driven by the pressure at the throat of the glottis. So, to evaluate the validity and accuracy of the simplified quasi-steady boundary-layer model, we compare the numeri-

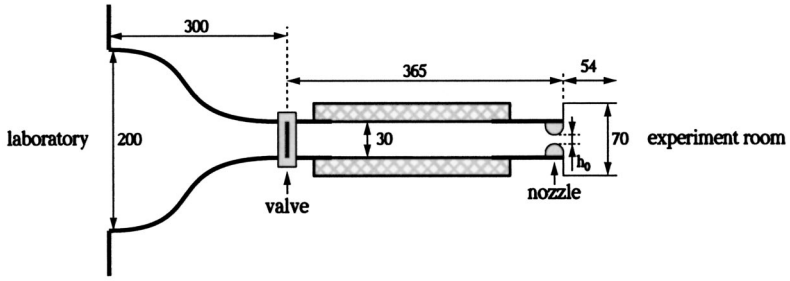


FIG. 3. Schematic representation of the experimental setup. Measures are in millimeters. h_0 is the (throat) height of the aperture (typically 1 or 3 mm). The shaded area represents the rubber foam padding on the inside of the upstream channel.

cally determined pressure at the throat to the theoretical prediction.

This paper has the following structure. We start by presenting some steady pressure measurements in a liplike round model and compare them to a simplified boundary-layer model. Next unsteady pressure measurements in both a liplike model and two diverging vocal-fold models are presented. Attention is paid to the various flow phenomena that are observed. Finally some results of numerical simulations of the flow through these models are presented. Both numerical and experimental results are compared to a quasi-steady boundary-layer model in order to study the applicability and accuracy of such a model in a description of the flow through the glottis.

II. EXPERIMENTS

A. Experimental setup

The experimental results that are presented in this paper have been obtained in the setup shown in Fig. 3. A sliding valve operated by a spring blade separates two large pressure reservoirs: laboratory (3500 m^3) and experiment room (75 m^3). A cylindrical pipe connects the inlet section to the glottis section. In order to damp mechanical and acoustic vibrations, the inner walls of this pipe are covered by a rubber foam (the shaded area in Fig. 3). The nozzle at the open pipe termination is a smooth constriction that is built out of either liplike round models or diffuser models that form a glottis-like channel. The brass blocks that form the geometry of the constriction are shown in Fig. 4. The block and hence the length of the glottal slit is 30 mm (i.e., the length in the third dimension not shown in the figure). All these blocks can be combined to form different geometries of the constriction but we will focus on the combinations that are presented in Fig. 4. The height h_0 representing the smallest aperture in the channel (the throat of the glottis) can be varied from 0.1 mm to 5 mm using calibrated spacers (metal rings) in the block mounting. Pressures are measured at two positions: the first position is 8 mm upstream of the start of the constriction in the side wall of the cylindrical pipe (piezo electrical transducer PCB 116A with a Kistler charge amplifier type 5011) and the second position is in the throat of the constriction at the smallest aperture. A piezo resistive pressure transducer (Kulite XCS062) is mounted in the blocks using a pressure tap with a diameter of 0.4 mm, as is shown in Fig. 5.

The liplike round model is built with two half-cylinders with a radius of 10 mm. The pressure tap is exactly in the middle of the block. The vocal-fold models consist of a cylindrical part, followed by a linearly diverging part and another cylindrical part (Scherer and Titze, 1983). The angle of divergence α is either 20° or 10°, as shown in Fig. 4. These models are the same ones as used by Pelorson *et al.* (1994) and are based on the typical vocal-fold movement discussed in the Introduction (see Fig. 1). The diffuser angles are chosen in such a way that when $h_0 \approx 3$ mm with $\alpha = 10°$ the flow would be in the stable-diffuser-flow region and with $\alpha = 20°$ the flow would be outside the stable-diffuser-flow region, according to data on diffuser performance at high Reynolds numbers (Re of the order 10^5) (Blevins, 1984). For the lip-

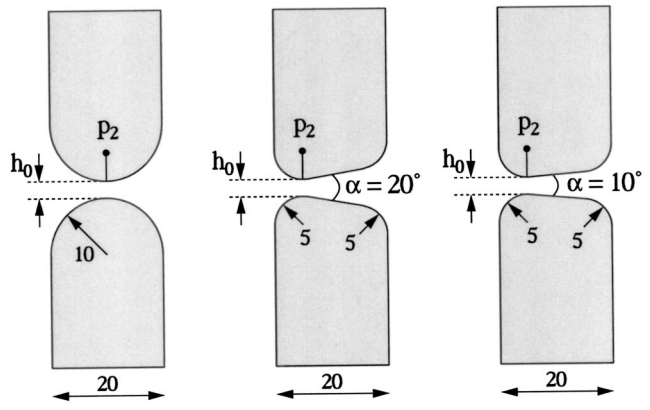


FIG. 4. Models that are studied: the model on the left is a liplike round model and the models on the right are diverging vocal-fold models. Measures are in millimeters. The arrows indicate the radii of curvature of the walls. The width of the blocks and hence the length of the glottal slit perpendicular to the flow direction is 30 mm.

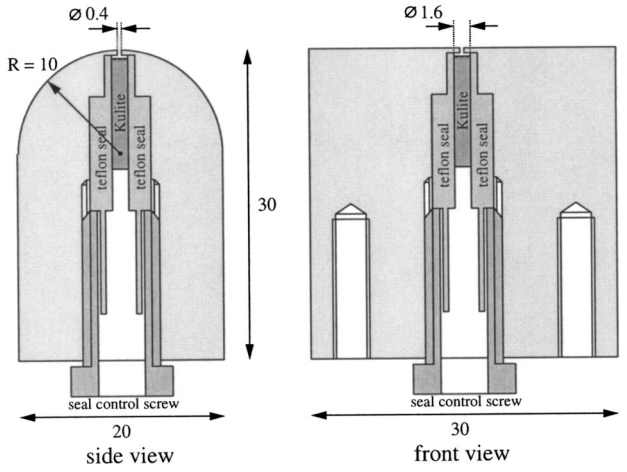


FIG. 5. Mounting of the Kulite pressure transducer in one of the brass blocks that form the constriction. A seal made of Teflon controlled by a screw is used to ensure a tight fit.

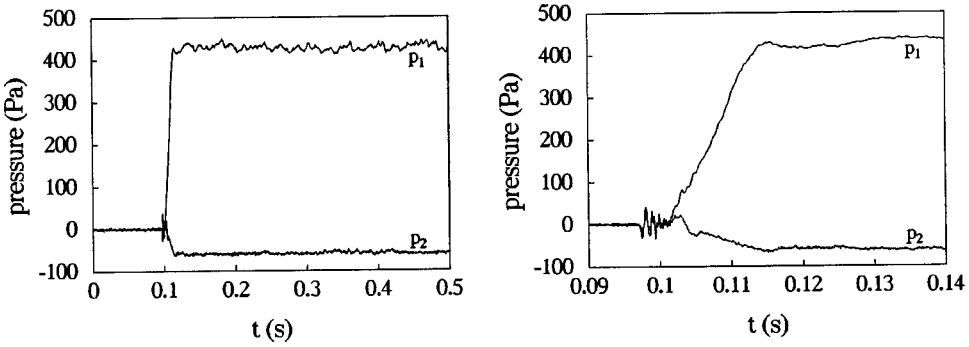


FIG. 6. Pressure measured in the pipe p_1 and in the throat p_2 for the liplike round geometry: $\Delta p = 430$ Pa and $h_0 = 0.99$ mm. The right graph is a close-up of the left graph, showing the transient behavior.

like model we will present in addition to the pressure data local velocity measurements obtained with a hot wire anemometer.

Using a laser detector system the opening of the valve is detected and the start of the measurement is triggered by this signal. During an experimental run pressures are measured and in some cases the velocity is measured simultaneously. The velocity is measured using a 4-μm-thick hot-wire in a constant-temperature anemometer. The signals are fed into a data acquisition system that is connected to a personal computer by means of a four-channel 12-bit ADC card (Keithley DAS-50). A typical experimental run lasts 500 ms while data are sampled at a frequency of 20 kHz.

In order to validate the accuracy of the pressure measurements, some experiments have been carried out using a straight channel with a smooth inlet and a sharp-edged outlet (Hofmans, 1998). Those tests confirmed the accuracy of the pressure measurements and the hot-wire anemometry, which was found to be of the order of 2%.

B. Results for the liplike round model

The liplike round model is studied for two reasons. First of all, the model can be considered a reference model for studying flow separation from a curved wall: because of the constant radius of curvature the separation point is usually not sensitive to external influences. The second reason is that this model is considered relevant for the study of the flow through the opening between the lips, which is important when considering plosives (Pelorson *et al.*, 1997) and brass instruments.

The experiments with the liplike models consist of pressure and velocity measurements. The pressure p_1 is measured in the pipe 8 mm upstream of the constriction while the pressure p_2 is measured at the smallest aperture (in the throat) of the constriction. The velocity is measured at various positions on the center line of the setup using the hot-wire anemometer. Experiments have been performed for two values of the throat height: $h_0 = 0.99$ mm and $h_0 = 3.36$ mm. It was found that the hot-wire probe disturbed the flow too much in the case of $h_0 = 0.99$ mm, so no velocity measurements have been done at the throat in this case. Since the results found for a throat height $h_0 = 0.99$ mm are very similar to results for $h_0 = 3.36$ mm, we mostly focus on results of the latter case.

In Fig. 6 a representative measurement is shown for the liplike models with a throat height $h_0 = 0.99$ mm. The final steady pressure drop Δp across the constriction is equal to 430 Pa. The opening time is approximately 15 ms, which, together with the length of the glottis model (20 mm) and the typical velocity $u_0 = \sqrt{2\Delta p/\rho}$, leads to a Strouhal number of 0.05. The Reynolds number is then approximately 1800 corresponding to loud speech. The left graph shows the pressure signals in the pipe p_1 upstream of the constriction and in the throat p_2 for a time range of 0.5 s. The right graph is a close-up of the left graph and shows the transient behavior in more detail in a time range of 0.05 s. The time axis is determined by the trigger for the measurement obtained by optical detection of the valve movement. In this and all subsequent graphs the trigger signal was generated at $t = 0.1$ s: by using the pretrigger capability of the ADC-card the measurement has already been recorded 0.1 s before the trigger is generated. The actual start of the flow is not determined since the optical detector setup is triggered by the valve movement and not by the flow, but it is reasonable to assume that this is close to the trigger point. In Fig. 7 two similar measurements with a throat height $h_0 = 3.36$ mm at a pressure drop of 290 Pa (Sr=0.05, Re=4900) and 690 Pa (Sr=0.03, Re=7600) are shown. Except for a few milliseconds before the trigger point the pressure in the pipe p_1 shows a smooth increase from zero to a steady value Δp. The small but distinct oscillations around the trigger point (at 0.1 s) are caused by the opening of the valve and could not be avoided in our setup. They also occur at zero pressure difference ($\Delta p = 0$). Hence we expect these oscillations to be due to acoustic waves generated by lateral movement of the valve before it actually opens. Using the optical detector setup the speed of the valve during opening is estimated to be in the range of 1.5 to 2 m/s. For an opening of 2 cm in the valve this corresponds to an opening time of the order of 10^{-2} s.

In Figs. 6 and 7 the pressure p_2 in the throat exhibits a particular behavior. The first few milliseconds of the experiments the pressure p_2 in the throat is rising proportionally to the pressure p_1 upstream of the glottis. This is due to the initial flow that is like a potential flow: boundary layers are still very thin and flow separation does not yet occur. The bulk flow is inviscid so velocity and pressure are related by Bernoulli's equation:

$$\rho \frac{\partial \phi_1}{\partial t} + \frac{1}{2}\rho u_1^2 + p_1 = \rho \frac{\partial \phi_2}{\partial t} + \frac{1}{2}\rho u_2^2 + p_2, \tag{1}$$

in which ρ is the density of air, u is the velocity, and ϕ is the velocity potential. The main contribution to the pressure drop

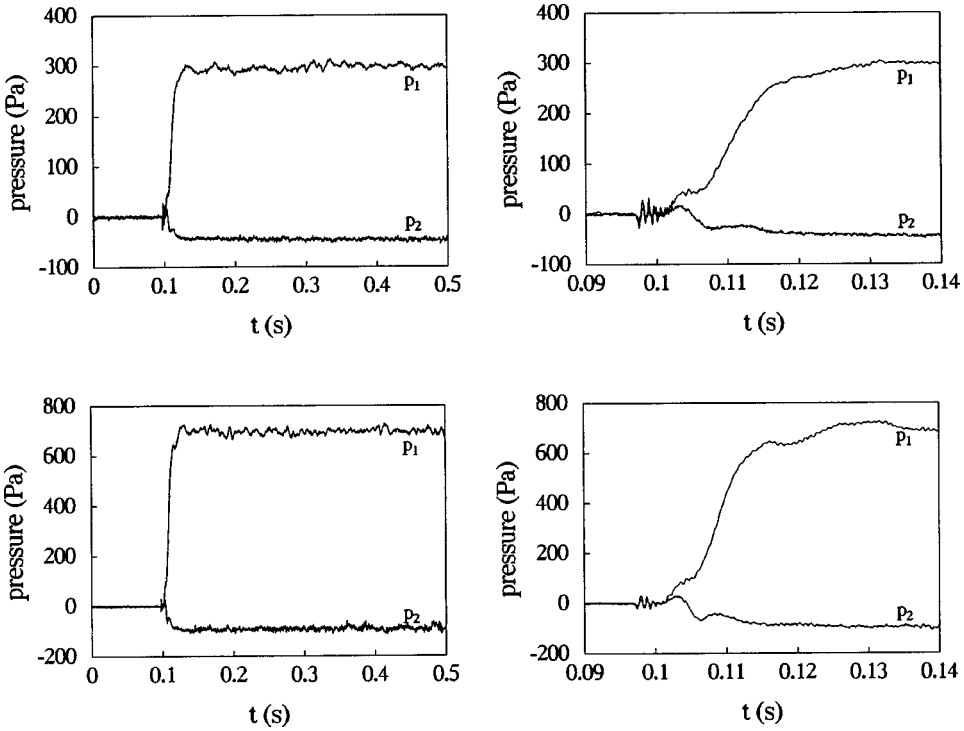


FIG. 7. Pressure measured in the pipe p_1 and in the throat p_2 for the liplike geometry: top graphs $\Delta p = 290$ Pa; bottom graphs $\Delta p = 690$ Pa, $h_0 = 3.36$ mm. The right graph is a close-up of the left graph, showing the transient behavior.

across the constriction is given by the inertial effects (the $1/2\rho u^2$-terms are negligible because the velocity is still very small). This implies that the pressure in the throat has a value that is somewhere between the pressure in the experiment room (by definition equal to zero) and the pressure in the pipe p_1. Typically p_2 is equal to 0.5 p_1 for the time interval $0.100\,\mathrm{s} < t < 0.105\,\mathrm{s}$, because the constriction is symmetric with respect to the throat and the main contribution to the inertial terms originates from this region. After the initial stage the flow separates from the walls of the constriction forming a jet. Finally, a steady situation is reached. If we assume the pressure in the jet to be equal to the pressure in the experiment room and if we assume the height of the jet h_s larger than the throat height h_0, then the pressure in the throat is lower than the pressure in the experiment room, since the velocity in the throat is higher than the velocity in the jet. This can be illustrated by inserting the one-dimensional equation of mass conservation $[u(x)h(x) = \Phi]$, thus neglecting boundary-layer effects, into the steady Bernoulli equation. Here $h(x)$ is the height of the channel and $\rho\Phi$ is the two-dimensional mass flux. This results in the following equation relating the pressure to the height of the channel:

$$\frac{p(x)}{\rho} + \frac{1}{2}\left(\frac{\Phi}{h(x)}\right)^2 = \frac{1}{2}\left(\frac{\Phi}{h_s}\right)^2, \tag{2}$$

in which h_s is the height of the channel at the separation point. By inserting p_1, pipe height h_p, p_2, and h_0 the following relationship between the pressure ratio p_2/p_1 and the various channel heights is obtained:

$$\frac{p_2}{p_1} = \frac{1-(h_s/h_0)^2}{1-(h_s/h_p)^2}, \tag{3}$$

which for $h_p \gg h_s$ reduces to

$$\frac{h_s}{h_0} = \sqrt{1-\frac{p_2}{p_1}}. \tag{4}$$

Therefore the pressure in the throat would be an indication of the jet width h_s and consequently of the position of the separation point. As can be observed in Figs. 6 and 7, p_2 is indeed less than zero in the steady limit.

Although the experiments for both values of h_0 look very similar, a few small but distinct differences can be observed. The experiment with $h_0 = 3.36$ mm show an oscillatory behavior in the pressures p_1 and p_2 during the first milliseconds after the trigger point. This oscillation is less pronounced in the experiments with $h_0 = 0.99$ mm and has a lower frequency. Furthermore, the experiments with $h_0 = 0.99$ mm seem to reach the steady state in a more straightforward way. We expect that this behavior is related to an acoustical resonance of the pipe system, which is not fully damped by the sound absorbing material in the upstream pipe segment (see Fig. 3). A more narrow slit can be associated with larger acoustical dissipation. The oscillation frequency is of the order of magnitude of the lowest acoustical mode of an open pipe segment of the length of the main pipe.

By integrating Bernoulli's equation the velocity in the throat can be computed based on the pressure measurements p_1 and p_2. For this purpose Bernoulli's equation is rewritten in the following form:

$$\frac{\partial(\phi_2 - \phi_1)}{\partial t} = \frac{1}{2}(u_1^2 - u_2^2) + \frac{p_1 - p_2}{\rho}, \tag{5}$$

which, using the definition of ϕ, can be written as

$$L_{\mathrm{eff}}\frac{\partial u_2}{\partial t} = \frac{1}{2}(u_1^2 - u_2^2) + \frac{p_1 - p_2}{\rho}, \tag{6}$$

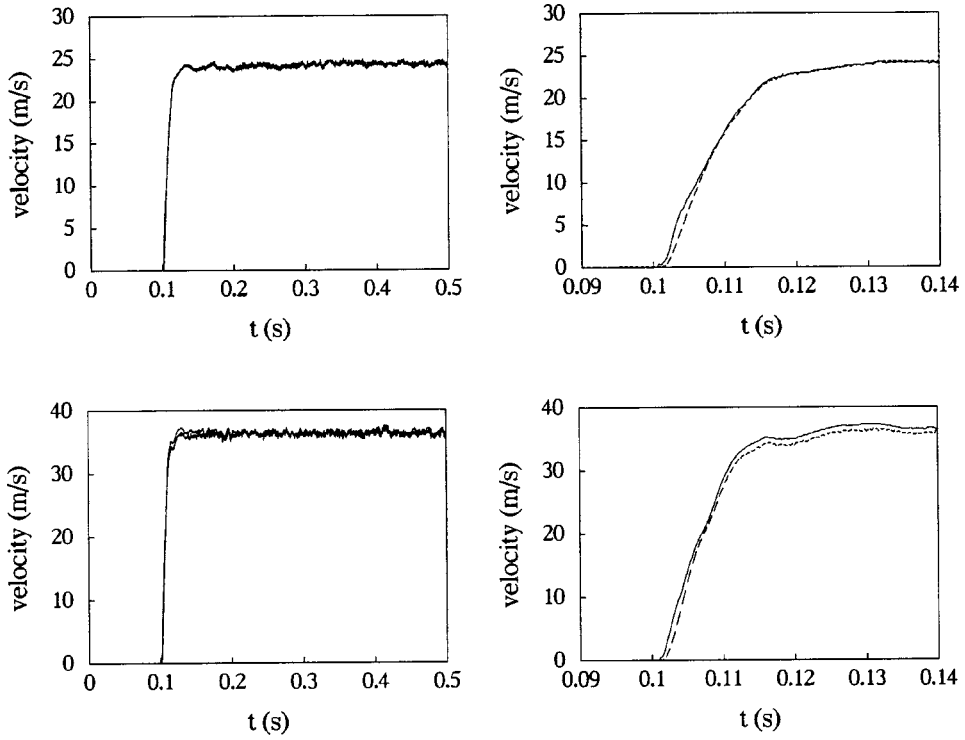


FIG. 8. Velocity measured in the throat (dashed line) and the velocity calculated in the throat by integration of the unsteady Bernoulli equation using the measured pressures p_1 and p_2 (solid line). Experiments performed on the liplike geometry: top graphs $\Delta p = 290$ Pa; bottom graphs $\Delta p = 690$ Pa, $h_0 = 3.36$ mm. The right graph is a close-up of the left graph, showing the transient behavior.

in which an effective length L_{eff} has been introduced. It is defined by

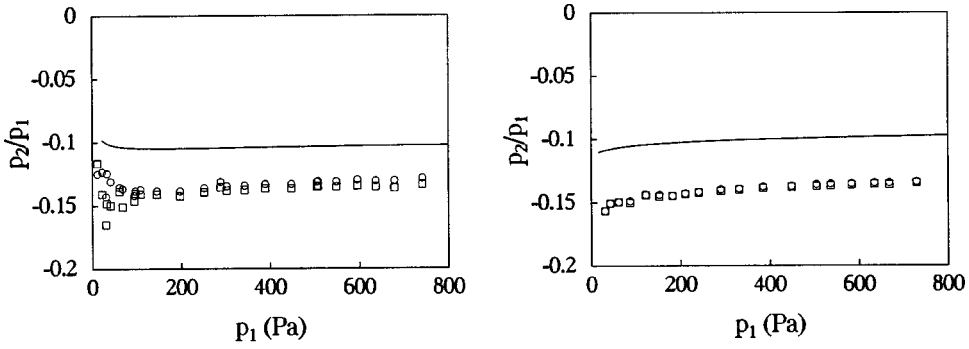
$$L_{\text{eff}} = \int_{x_1}^{x_2} \frac{h_0}{h(x)}\, dx, \tag{7}$$

where in the last step we assumed a uniform flow across the channel for any given x-position (quasi-one-dimensional flow).

In Fig. 8 the computed velocity (solid line) is compared to the velocity measured in the throat by means of the hot-wire anemometer (dashed line). The velocity that is computed is the velocity at the edge of the boundary layer, which does not have to agree with the velocity on the center line in a curved channel, because of the effect of curved streamlines on the pressure. It has been found, however, that this effect is considerably reduced in the throat because the boundary-layer growth in this region tends to cancel this effect. Furthermore, since flow separation takes place close to the throat, the curvature effect is also reduced (even neglecting boundary-layer growth). For this reason an assumption of uniform flow through the throat is quite reasonable. The pressure measurements of these experiments are shown in Fig. 7. A good overall agreement is found between hot-wire measurement and calculation of the velocity in the throat based on pressure measurements. This confirms the accuracy of the pressure measurements. However, in the initial stage of the experiments ($0.1 < t < 0.105$) the hot-wire measurements are delayed with respect to the computed velocity profiles, similar to the results found for the sharp edge nozzle configuration (Hofmans, 1998). This is expected to be due to a poor dynamical response of the hot-wire at low velocities.

Since one of the aims of this paper is to determine the validity of simplified models, results of the boundary-layer theory are compared to steady pressure measurements in the liplike models. The theoretical results are based on Pohlhausen's method (Pohlhausen, 1921) using a third-order polynomial to describe the velocity profile in the boundary layer and then solving the steady von Kármán equation as done previously by Pelorson *et al.* (1994). There are unfortunately some printing errors in the formulas provided by Pelorson *et al.* Correct formulas are provided in the Appendix. The steady pressures are measured by means of Betz water manometers with an accuracy of 0.05 mm H_2O ($\approx$0.5 Pa). Results are shown in Fig. 9. The markers represent measurements in the top and bottom walls of the constriction. The solid line represents the theoretical result.

The boundary-layer theory predicts a too high value of p_2/p_1. This might be the result of the assumptions that are made. In the theoretical model, the boundary layer is calculated up until the separation point. At the separation point the pressure is assumed to be equal to the external pressure (quasi-steady uniform velocity jet model). It is also possible to continue the boundary-layer calculation beyond the separation point. This leads to a lower prediction of p_2/p_1. However, these calculations do not converge to a constant pressure value as we increase the calculation domain and they are therefore not reliable. The third-order polynomial does not describe the flow accurately (far) beyond the separation point. These results may indicate that the assumption of constant pressure in the jet is not valid inside the constriction and that the reference point for the pressure in the boundary-layer model is not correctly chosen. Alternatively this might just be due to a poor prediction of the separation point by the boundary-layer model. This last hypothesis will be eliminated by comparison of boundary-layer theory with numerical calculations presented in Sec. III.



(a) $h_0 = 0.99$ mm

(b) $h_0 = 3.36$ mm

FIG. 9. Comparing boundary-layer predictions (solid line) with steady pressure measurements at the throat in the top wall (○) and the bottom wall (□) of the liplike models.

C. Results for the 20°-diverging model

Models that are more specific for the study of the flow through the glottis are the two diverging scale models that are shown in Fig. 4. In this section we present results for the diffuser model with a total angle of divergence α of 20°. The flow through this model is investigated by means of unsteady pressure measurements at two positions: one in the pipe 8 mm upstream of the constriction and the other in the throat of the constriction. Measurements are done with two values for the height of the throat: $h_0 = 1.12$ mm and $h_0 = 3.50$ mm.

In Fig. 10 pressure measurements are presented for $h_0 = 1.12$ mm and two values of the steady pressure drop: $\Delta p = 301$ Pa (Re$=1.5\times10^3$) and $\Delta p = 627$ Pa (Re$=2\times10^3$). The right graphs show a close-up of the left graphs, focusing on the transient behavior. In fact each graph consists of two experimental results. Repeating the experiment several times we found that two distinctly different time histories for the pressure p_2 were obtained. As can be observed in Fig. 10, initially the two time histories collapse onto one curve and only after a certain time the two curves start to deviate, leading to two different steady-pressure values for p_2. This behavior is explained by an asymmetric flow due to the so-called Coanda effect (Tritton, 1988). The Coanda effect is due to viscous entrainment of the air that is caught between the jet and the walls of the channel. The symmetric jet becomes meta-stable and a small perturbation results in an adherence of the jet to either the top or bottom wall of the channel. Since this phenomenon can be triggered by a small asymmetry in the flow, both states are possible in a symmetric setup. The pressure signal p_1 is very similar to the signals found in the liplike model. Also the initial increase in p_2—corresponding to an unsteady potential flow—can be observed in this case. However, here the similarities with the results of the liplike model end and a different behavior of p_2 is observed. Note the rather strong downward peak in p_2 just before the steady limit is reached.

In Fig. 11 equivalent pressure measurements are presented for $h_0 = 3.50$ mm and two values of the steady pressure drop: $\Delta p = 268$ Pa (Re$=4\times10^3$) and $\Delta p = 528$ Pa (Re

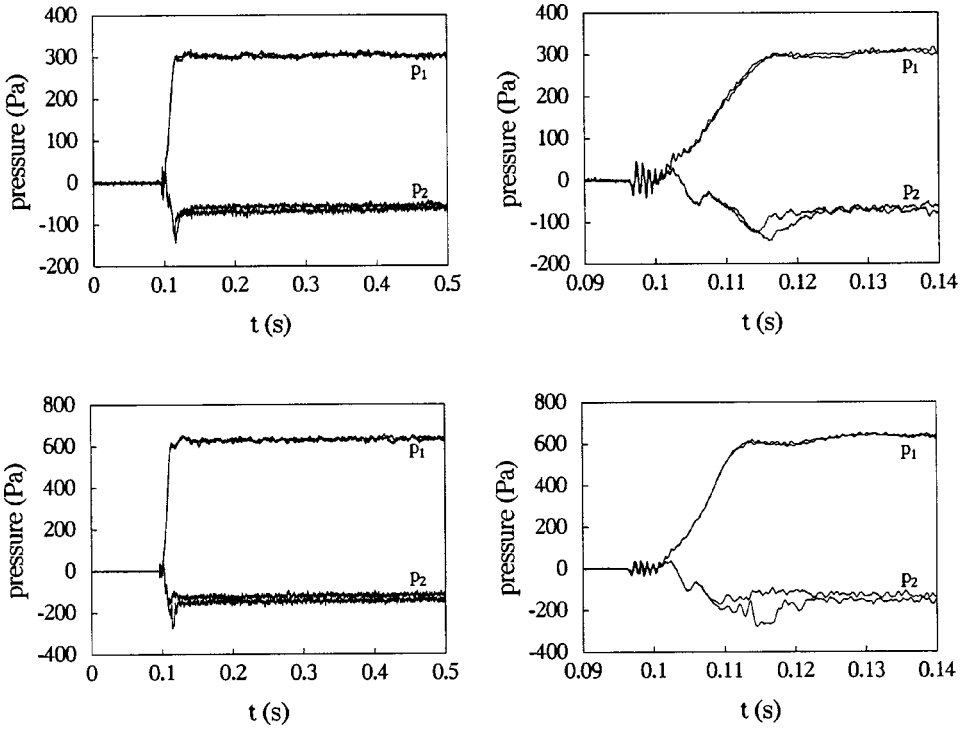


FIG. 10. Pressure measured in the pipe p_1 and in the throat p_2 for the diffuser geometry with α=20°: top graphs $\Delta p = 301$ Pa; bottom graphs $\Delta p = 627$ Pa, $h_0 = 1.12$ mm. The right graph is a close-up of the left graph, showing the transient behavior. In each graph two experimental runs are shown.

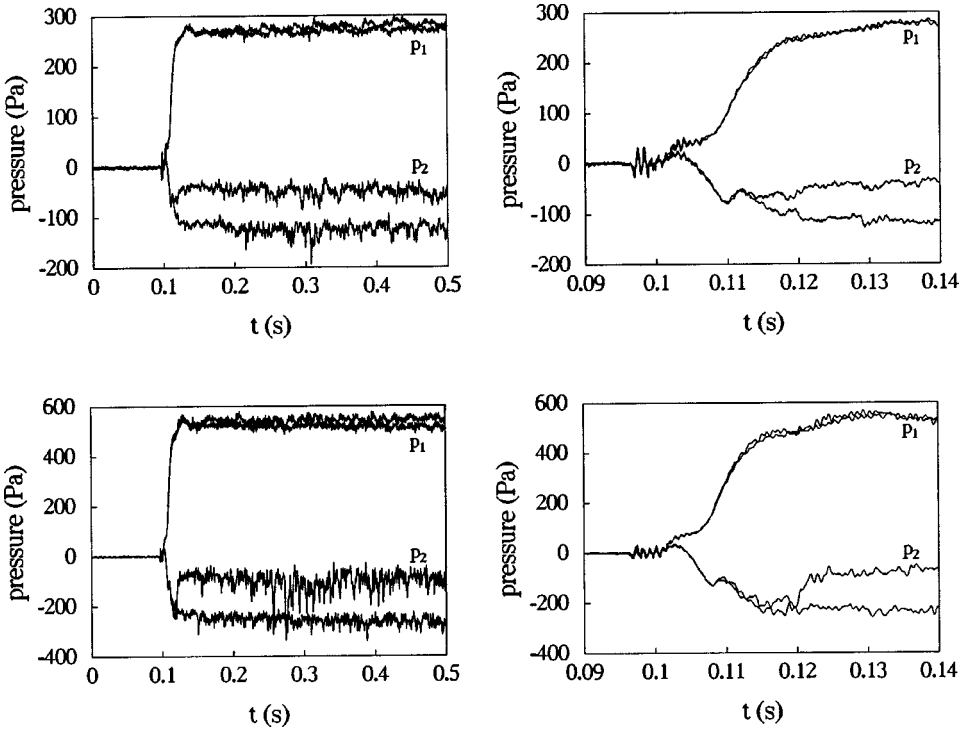
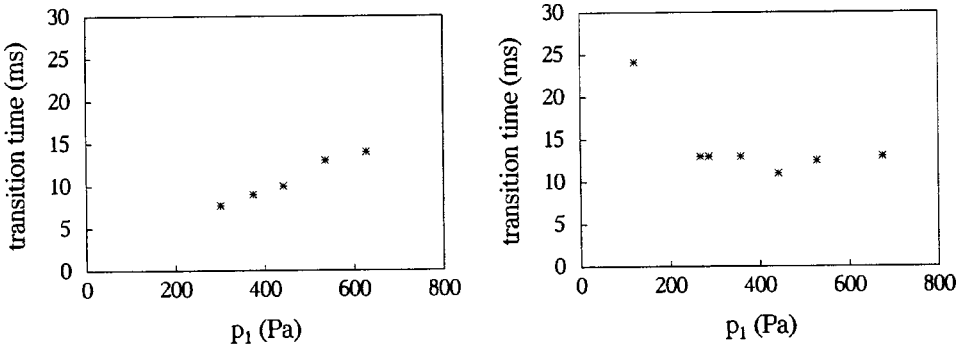


FIG. 11. Pressure measured in the pipe p_1 and in the throat p_2 for the diffuser geometry with $\alpha=20°$: top graphs $\Delta p=268$ Pa; bottom graphs $\Delta p=528$ Pa, $h_0=3.50$ mm. The right graph is a close-up of the left graph, showing the transient behavior. In each graph two experimental runs are shown.

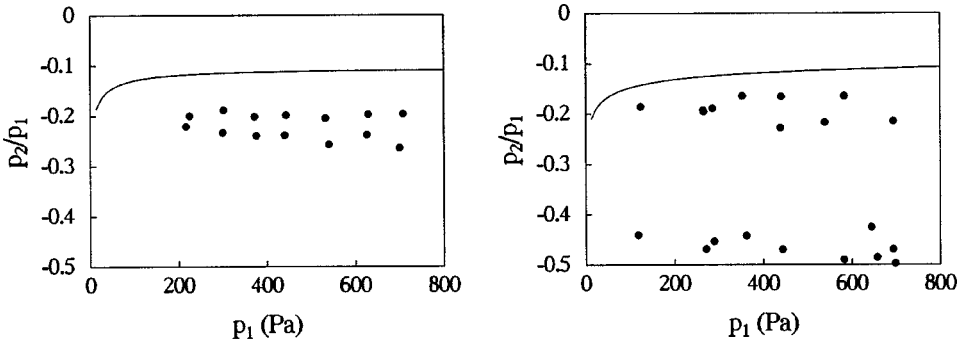
$=6\times10^3$). The Coanda effect is also observed in this configuration and the behavior is very similar to the previous case, except that the bifurcation of the two states is much more prominent. The presence of the Coanda effect was confirmed by flow visualizations and by simultaneous measurements at both sides of the throat using less accurate pressure transducers. Pelorson and Hirschberg (1997) also presented accurate simultaneous pressure measurements that confirm the occurrence of a Coanda effect. Note also the increased level of the fluctuations on the pressure signals which might indicate the onset of turbulent flow. In general it was found that the flow was very unstable in this configuration.

Both Figs. 10 and 11 illustrate the fact that the establishment of the Coanda effect takes time. In order to quantify this claim, the time at which the two time histories of p_2 start to deviate is measured as a function of the steady pressure drop. The time is taken with respect to the trigger point, since this is a good indication of the actual start of the flow. The results are presented in Fig. 12. Similar results have been presented by Pelorson and Hirschberg (1997). The results for $h_0=1.12$ mm seem to exhibit a linear relationship with the pressure drop, although data at low pressures ($p_1<300$ Pa, $\mathrm{Re}<1.5\times10^3$) are lacking. On the other hand, the results for $h_0=3.50$ mm do not exhibit this and in fact for a large range of p_1 ($250\,\mathrm{Pa}<p_1<700\,\mathrm{Pa}$, $4\times10^3<\mathrm{Re}<8\times10^3$) the transition time remains unchanged at approximately 13 ms (translated to physiological conditions 1.5 ms). Below 250 Pa the transition time increases strongly to 25 ms (translated to physiological conditions 2.8 ms). The transient in our glottis model lasts approximately 20 ms, while the related closing or opening time of the glottis oscillation is approximately 2 ms for typical male voiced sound production at 100 Hz (see Fig. 2). So the significance of the Coanda effect is not yet clearly established. On the other hand, in the real glottis the movement of the walls is an important factor and the consequences of this movement for the Coanda effect are yet to be determined but could be quite significant. In the closing phase of the glottis, the Coanda effect is expected to be reduced because the wall movement is equivalent to an injection of fluid at the wall. The opposite occurs



(a) $h_0 = 1.12$ mm (b) $h_0 = 3.50$ mm

FIG. 12. The time with respect to the trigger point at 0.1 s at which the flow starts to exhibit two states due to the Coanda effect for the diffuser model with $\alpha=20°$.



(a) $h_0 = 1.12$ mm

(b) $h_0 = 3.50$ mm

FIG. 13. Comparing boundary-layer predictions (solid line) with steady limit pressures obtained from the unsteady experiments (markers): vocal-fold model with $\alpha = 20°$.

during the opening phase. However, in the opening phase the glottis is convergent and the Coanda effect is expected to be only a minor effect.

In order to compare the experimental results with predictions of the boundary-layer theory, the steady pressure ratio p_2/p_1 is determined in the unsteady experiments. The theoretical results are again based on Pohlhausen's method (Pohlhausen, 1921) using a third-order polynomial to describe the velocity profile in the boundary layer. Results are shown in Fig. 13. For each value of p_1 two values of p_2 are found experimentally. The different values are due to the adherence of the flow to either the upper wall of the constriction that contains the pressure tap (lower results) or to the opposite wall (upper results). For $h_0 = 1.12$ mm the two sets of values differ by 25%, while for $h_0 = 3.50$ mm they differ by a factor of 2. Again the theoretical prediction results in a too high value of p_2 (i.e., not negative enough), although a comparison is not completely justified due to the assumption of a symmetric solution in the theoretical boundary-layer model. Similar to the results found in the liplike model, an increase in the pressure downstream of the separation point is neglected. Because of the asymmetry of the flow, no conclusions on the point of separation can be drawn based on the pressure measurements.

D. Results for the 10°-diverging model

In this section we present results for the diffuser model with a total angle of divergence α of 10°. The flow through this model is again investigated by means of unsteady pressure measurements at two positions: one in the pipe 8 mm upstream of the constriction and the other in the throat of the constriction. Measurements are done with two values for the height of the throat: $h_0 = 1.01$ mm and $h_0 = 3.39$ mm.

In Fig. 14 pressure measurements are presented for $h_0 = 1.01$ mm and three values of the steady pressure drop: $\Delta p = 224$ Pa (Re $= 1.2 \times 10^3$), $\Delta p = 432$ Pa (Re $= 1.8 \times 10^3$), and $\Delta p = 682$ Pa (Re $= 2.1 \times 10^3$). The right graphs show a close-up of the left graphs, focusing on the transient behavior. The pressure signal p_1 is very similar to the signals found in the previous two scale models. Also the initial increase in p_2 can be observed in this case. However, again here the similarities with previous results end and different behavior of p_2 is observed. With increasing pressure an increasing level of fluctuations can be observed in p_2. Especially large fluctuations are observed for $\Delta p = 682$ Pa. This behavior is similar to the behavior observed in the 20°-diverging model with $h_0 = 3.50$ mm. However, p_1 is much more stable in this case and only one value is found for p_2. This behavior is apparently not due to the Coanda effect and simultaneous pressure measurements at both sides of the glottis displayed symmetric behavior (Pelorson *et al.*, 1995). A possible explanation is a shift of the very unstable separation point due to the transition from laminar flow to turbulent flow. Turbulence enhances strongly the flow entrainment by the jet, thus creating a low pressure region between the wall and the jet downstream of the separation point which pushes the separation point downstream. This is a symmetrical effect which occurs at both sides of the jet. This effect is even more prominent in the results obtained with $h_0 = 3.39$ mm as presented in Fig. 15. Three sets of pressure measurements are presented for $\Delta p = 161$ Pa (Re $= 3 \times 10^3$), 407 Pa (Re $= 5 \times 10^3$), and 654 Pa (Re $= 6 \times 10^3$). The top and center graphs show a distinctive, abrupt transition from a temporary stable solution to a new stable solution, while in the bottom graph this transition is already occurring during the initial transient. The difference in the level of the fluctuations before and after the transition indicates a transition from a laminar jet flow to a turbulent jet flow. The jet turbulence is indeed clearly observed in the flow visualizations (Pelorson *et al.*, 1994). Since a turbulent jet has a much stronger entrainment than a laminar flow, the net effect is to delay separation until the end of the diffuser when the diffuser angle is small. This would result in a much lower value for p_2. In fact we will show that the glottis model is acting like a well-designed diffuser for this configuration when turbulence has appeared in the jet flow. This behavior also agrees with the data presented in Blevins (1984) at much higher Reynolds numbers (Re of the order 10^5).

In Fig. 16 the steady pressure limits are compared to the theoretical prediction based on Pohlhausen's method (Pohlhausen, 1921) using a third-order polynomial to describe the velocity profile in the laminar boundary layers. In the case $h_0 = 1.01$ mm no abrupt transition is observed and therefore only one set of values is plotted in the left graph of Fig. 16. For low values of p_1 the agreement with the boundary-layer prediction is reasonable. An increasing deviation from the boundary-layer prediction is found with increasing value of p_1. In the case $h_0 = 3.39$ mm, a distinctive abrupt transition

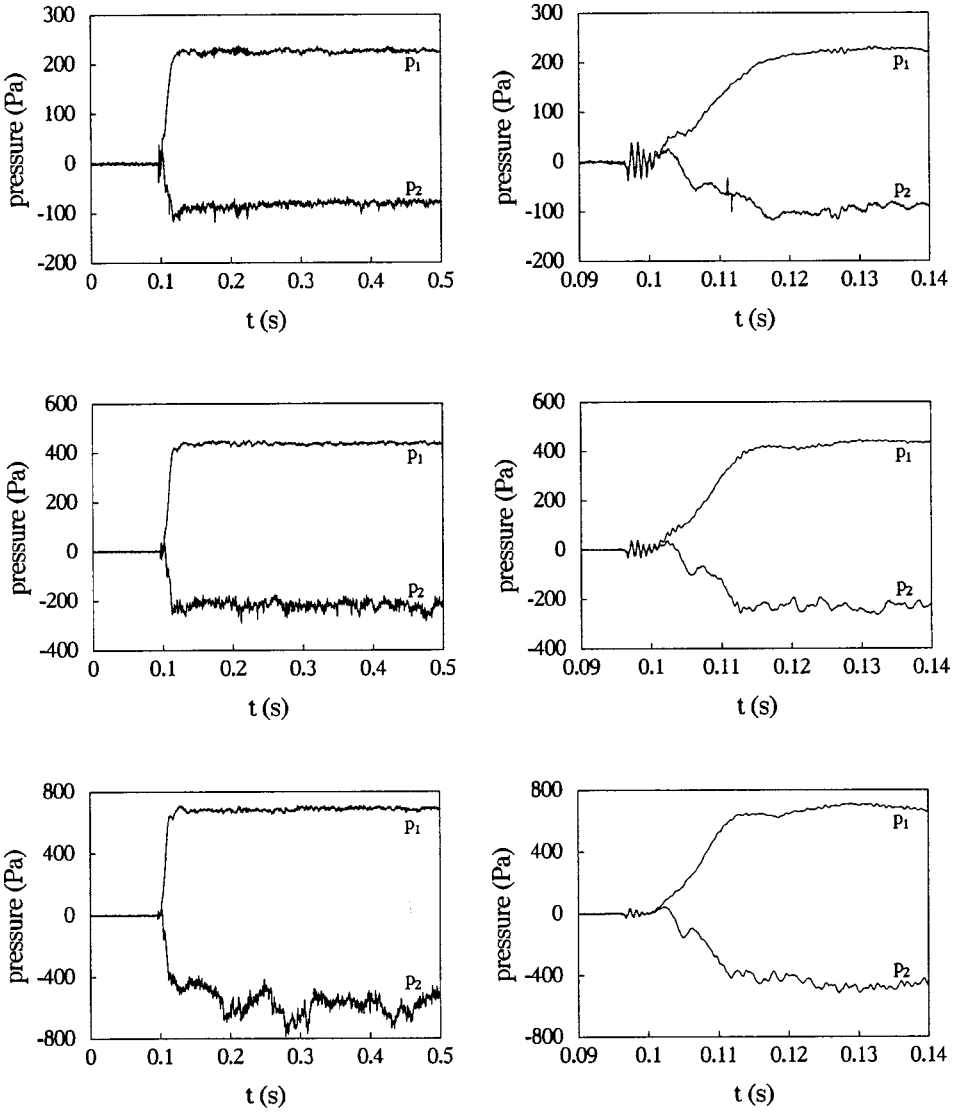


FIG. 14. Pressure measured in the pipe p_1 and in the throat p_2 for the diffuser geometry with $\alpha=10°$: top graphs $\Delta p=224$ Pa; center graphs $\Delta p=432$ Pa; bottom graphs $\Delta p=682$ Pa, $h_0=1.01$ mm. The right graph is a close-up of the left graph, showing the transient behavior.

is found for low values of p_1. In that case, two values for p_2 can be determined, one value before the transition and one value after the transition. For higher values of p_1 the transition occurs too early, so that it is not possible to determine a stable laminar value for p_2 before the transition. The values of p_2 before the transition at low values of p_1 agree quite well with the theoretical prediction. The turbulent results p_2 of course do not agree with the prediction of the laminar boundary-layer theory.

Using Eq. (4) the height h_s at the separation point can be estimated from the pressure ratio. In Fig. 17 the results are plotted. Clearly visible is the shift of the separation point in the downstream direction due to turbulence. This means that the flow remains a further distance attached to the walls of the glottis model as the pressure difference in the experiments increases. For $h_0=3.39$ mm a limit is reached above $p_1=200$ Pa (Re=4500). The limit $h_s/h_0\approx 1.55$ in the right graph of Fig. 17 coincides with the value of h/h_0 at the end of the linearly diverging part of the constriction. In that case we do not expect a significant pressure recovery in the free jet.

In Fig. 18 the moment at which the transition occurs is plotted as a function of the pressure drop p_1 across the constriction. This figure illustrates that like the Coanda effect, turbulence needs a long time before it is established. For a typical Reynolds number relevant in speech ($\mathrm{Re}_h=3000$), the time delay in our experiments is 100 ms which corresponds to 11 ms under physiological conditions. For normal male voiced sound the oscillation period is of the order of 10 ms (see Fig. 2) which corresponds to 6 ms open phase. So, similar to the Coanda effect, in a pulselike flow with a duration of the order of 10 ms this effect might not be important in normal speech. However, Fig. 15 shows that at high pressures, corresponding to shouting, the pressure in the throat of the glottis never reaches a steady laminar value before turbulence sets in completely. It is further expected that such effects are sensitive to the movement of the walls and hence this can induce a hysteresis in the flow separation behavior as a function of h_0.

III. NUMERICAL SIMULATION

A. Method

The incompressible two-dimensional Navier–Stokes equations are solved using the so-called viscous vortex-blob

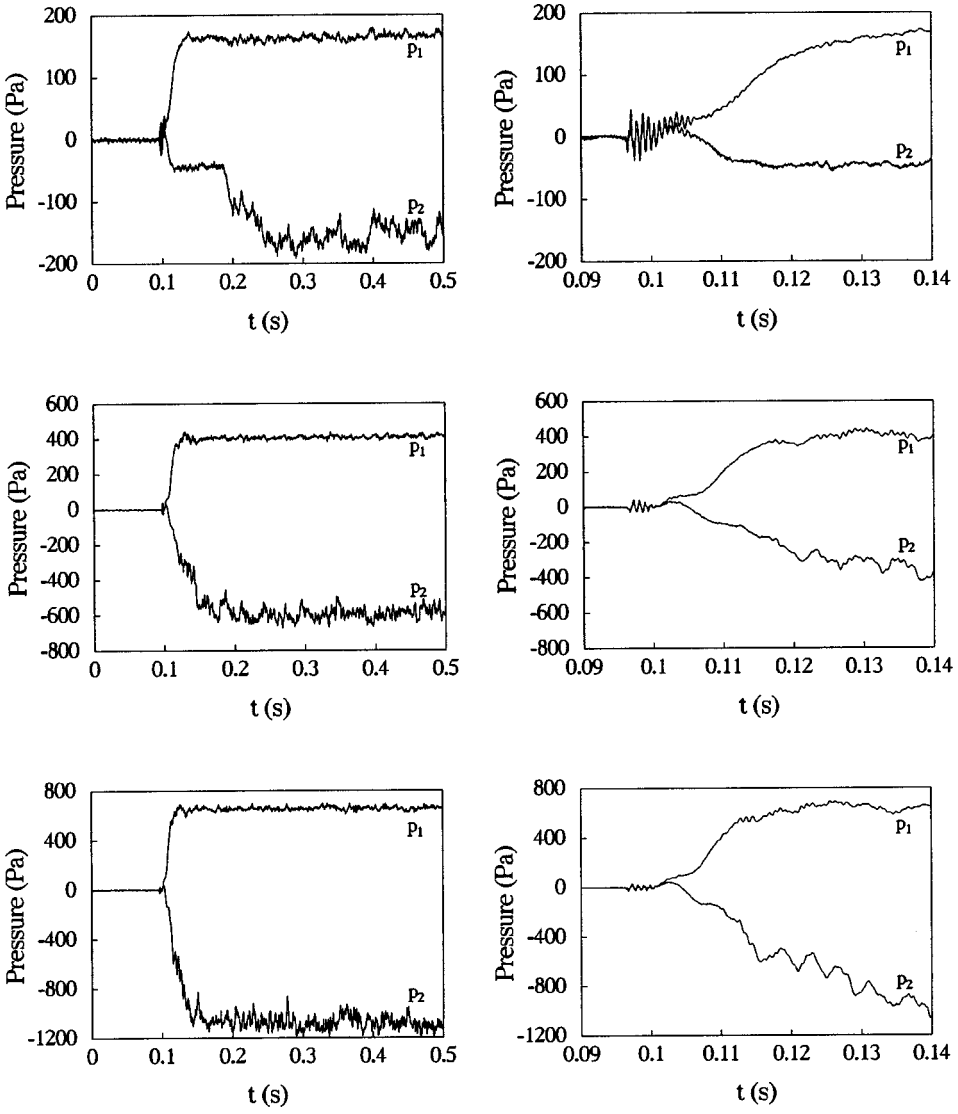
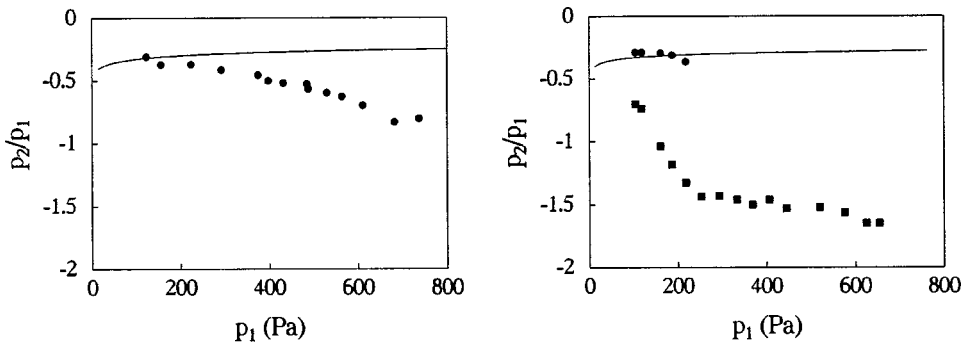


FIG. 15. As in Fig. 14, with $\alpha=10°$: top graphs $\Delta p=161$ Pa; center graphs $\Delta p=407$ Pa; bottom graphs $\Delta p=654$ Pa, $h_0=3.39$ mm.

method (Graziani *et al.*, 1995; Ranucci, 1995). The method is based on a desingularized point vortex method. The solution is obtained in the form of the vorticity distribution ω by solving the vorticity-transport equation:

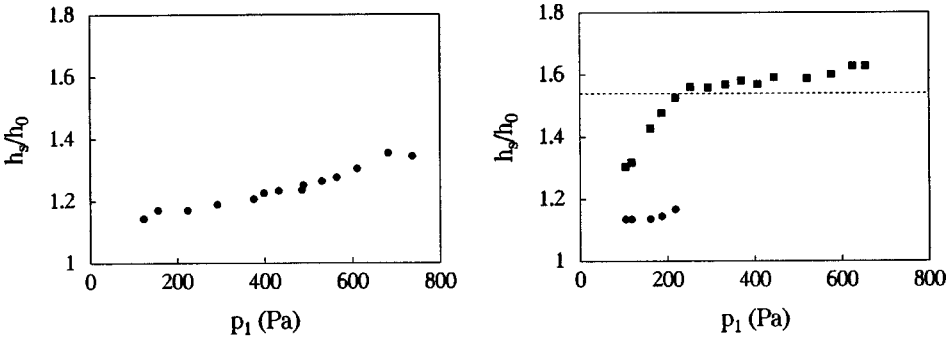
$$\frac{\partial \omega}{\partial t}+\mathbf{u}\cdot\nabla\omega=\frac{1}{\mathrm{Re}}\nabla^2\omega, \tag{8}$$

where $\mathbf{u}$ is the local velocity field and Re is the Reynolds number. For an appropriate treatment of both the convective and the diffusive time scale as well as for the accurate approximation of the nonlinear term, the equation has been split into a ''Euler step'' and a ''Stokes step'' according to the Chorin–Marsden product formula (Chorin *et al.*, 1978). The first step is governed by the inviscid-flow equation stating



(a) $h_0 = 1.01$ mm (b) $h_0 = 3.39$ mm

FIG. 16. Comparing boundary-layer predictions (solid line) with steady limit pressures obtained from the unsteady experiments: ● before transition; ■ after transition. No transition occurred in the experiments with $h_0=1.01$ mm. Diffuser model with $\alpha=10°$.



(a) $h_0 = 1.01$ mm

(b) $h_0 = 3.39$ mm

FIG. 17. The ratio of the channel height at the separation point and the throat height h_s/h_0 as computed from p_2/p_1 using Eq. (4): ● before transition; ■ after transition. The dashed line represent the height at the end of the linearly diverging section. Diffuser model with $\alpha=10°$.

that the material identity of the vortical particles is advected by the velocity field:

$$\frac{D\omega}{Dt}=0. \tag{9}$$

In the second step, Stokes' equation in vorticity form in two dimensions becomes

$$\frac{\partial\omega}{\partial t}=\frac{1}{\mathrm{Re}}\nabla^2\omega. \tag{10}$$

The fractional-step scheme provides the most appropriate and convenient solution for each substep. The vorticity field defined on the computational body-fitted grid with mesh size h is approximated as

$$\omega(\mathbf{x},t^n)=\sum_{j=1}^{N}\Gamma(\mathbf{x}_j,t^n)\,\delta(\mathbf{x}-\mathbf{x}_j), \tag{11}$$

where $\Gamma(\mathbf{x}_j,t_n)=\omega(\mathbf{x}_j,t_n)h^2$ is the circulation at the grid point $\mathbf{x}_j$ and time t_n and $\delta(\mathbf{x}-\mathbf{x}_j)$ is the Dirac delta function. Since the interaction between two point vortices diverges as the point vortices approach each other, a desingularization of this interaction is appropriate as has been first discussed by Chorin and Bernard (1973) and later by others such as Beale and Majda (1985). In our method the point-vortex interaction is desingularized by the higher-order kernel of Lucquin-Descreux (1987). Further details on the numerical method can be found in Hofmans (1998).

B. Input for the simulations

Using the viscous vortex-blob method, four simulations have been done with the diverging vocal-fold models at moderate Reynolds number (of the order 10^3): $\alpha=20°$, $h_0=1.05$ mm, and $h_0=3.35$ mm; $\alpha=10°$, $h_0=1.05$ mm, and $h_0=3.35$ mm. The numerical method yields a solution of the two-dimensional incompressible Navier–Stokes equations and needs an imposed inflow velocity (u_{in}) as input. The steady-state value of the inflow velocity is used as the reference velocity u_{ref}. This input is obtained from experimental pressure measurements (as presented in the previous paragraphs). First a ninth-order polynomial is fitted to the pressure difference (p_1-p_2) between pipe and throat as a function of time. Next this smooth fitted pressure profile is integrated in time using the unsteady Bernoulli equation to find the inflow velocity as a function of time. This method is more practical than a polynomial fit to the velocity profile because this usually requires at least two separate fits. In Fig. 19 an example is shown of the result of this procedure. In the left graph the pressure difference (p_1-p_2) is plotted together with the fitted polynomial. In the right graph the resulting scaled inflow velocity is plotted as a function of time. Time equal to 0 is defined to be the start of the simulation.

The computational domain for the diffuser models is shown in Fig. 20. The inflow-velocity (u_{in}) is uniformly imposed on the left inflow boundary. The height of the inflow channel is related to the radius of curvature r and the throat height h_0. It is defined as the reference length l_{ref} in the simulation $l_{\mathrm{ref}}=4r+h_0$. On the right the outflow boundary is a far-field semi-circular domain (at a distance of 27 l_{ref}). On this boundary a radial outflow is assumed and the outflow velocity is determined from the inflow velocity by the continuity equation. Since the geometry that we are interested in is symmetric with respect to its center line, the computation is restricted to the upper half of the domain, hence the use of a symmetry line as a lower boundary and only half a glottis as the upper boundary. This also reduces the computational time and memory requirements by approximately a factor of 2. The boundary is discretized by a set of panels. A densification of panels is applied in the region around the glottis and on the symmetry line. A typical run requires 1600 to 1900 panels to build the geometry and uses 50 000 to 100 000 point vortices to discretize the vorticity field. The region in which the full (viscous) Navier–Stokes equations are solved is restricted to the viscous domain. Outside this

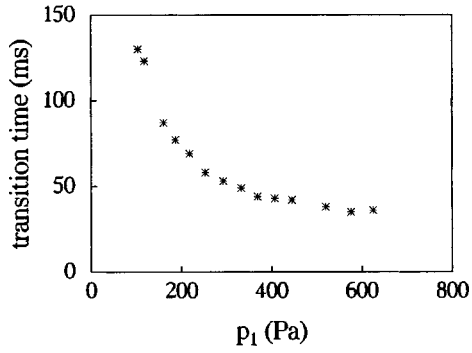


FIG. 18. The time with respect to the trigger point at 0.1 s at which the flow starts to change significantly due to the onset of turbulence. Diffuser model with $\alpha=10°$.

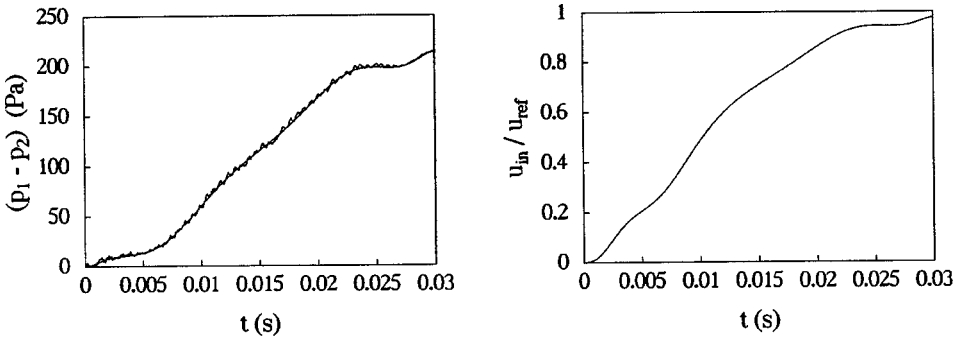


FIG. 19. Input for a numerical simulation. The scaled inflow u_{in}/u_{ref} (right graph) for the numerical simulation is obtained from the pressure difference (p_1-p_2) (left graph) by fitting a ninth-order polynomial (left graph) and integrating the unsteady Bernoulli equation, steady-state inflow velocity $u_{ref}=2.609$ m/s.

domain only the inviscid Euler equations are solved. The viscous domain includes the whole jet-region and the region of flow separation (which is in fact the whole constriction area). In the inflow domain and in the upper region of the outflow channel viscous effects are neglected since they have negligible influence on the results.

C. Results

The inflow boundary condition has been obtained from experimentally measured pressure differences (p_1-p_2) in the four geometries. The experimental pressures are shown in Fig. 21 represented by the thin lines. The pressure p_1 was measured 8 mm upstream of the constriction while the pressure p_2 was measured in the throat of the constriction. Also shown are the numerical results for p_1 and p_2 by means of the thick lines. Although the experimental pressure difference (p_1-p_2) was used to determine the inflow velocity, this does not impose individual values of either p_1 or p_2. The results in Fig. 21 demonstrate that the numerical method yields very reasonable results for the pressures p_1 and p_2. Note that the very unstable behavior that was found experimentally for the case $\alpha=20°$ and $h_0=3.50$ mm (see Fig. 11) seems also to be present in the numerical result for $\alpha=20°$ and $h_0=3.35$ mm, shown in the top right graph of Fig. 21. An explanation for this unstable behavior is suggested by a close-up of the numerical vorticity in the diffuser region of the glottis shown in Fig. 22. As these plots show, the interaction between the jet and the walls of the diffuser is very complex. After 0.02 s, bursts of vorticity leave the walls at regular intervals in time. Oscillatory pressure fluctuations in experiments could well be associated with this behavior.

In order to study the behavior of the separation point in these flows, the calculated height h_s of the channel at the separation point is plotted versus time in Fig. 23. As references, the heights that are associated with the starting point and the end point of the linear divergent part of the diffuser section are indicated by the dashed horizontal lines. The definition of a separation point is not obvious in an unsteady flow. We used the definition of the separation point for steady flows: $\partial u/\partial y=0$ on the wall, which is equivalent to $\omega_{wall}=0$. This choice was made because it makes a comparison to a quasi-steady boundary-layer model straightforward.

The behavior visible in the top two graphs for $\alpha=20°$ is very different from that in the bottom two graphs for $\alpha=10°$. First of all, for $\alpha=20°$ flow separation starts somewhere halfway in the constriction and moves very rapidly to the throat of the constriction. A steady value just downstream of the constriction on the cylindrical part is reached after 10 ms. In the $\alpha=10°$ case, flow separation starts at the end of the constriction on the cylindrical part. It then moves rapidly to a stable point somewhere on the diffuser part of the vocal-fold model. The model is acting like a diffuser with marginal flow separation [also called stall (Blevins, 1984)], and a small perturbation can have a significant influence on the separation point. The sensitivity of the separation point to small perturbations observed in those laminar simulations support the hypothesis that a transition from a laminar to a turbulent flow results in a sudden change of the position of the sepa-

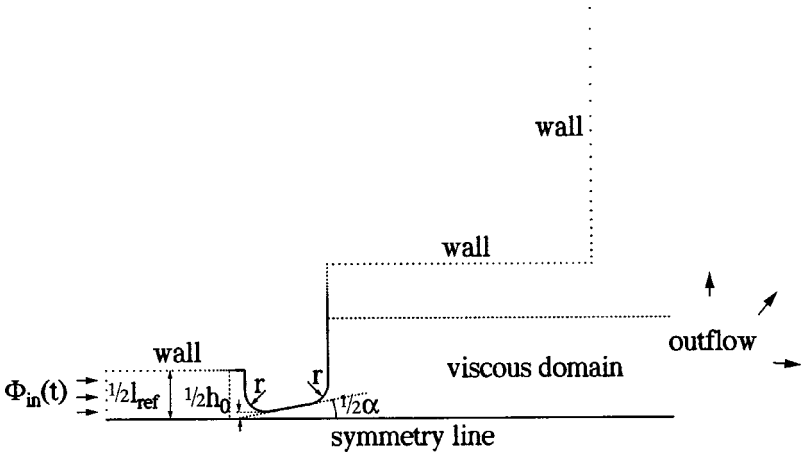


FIG. 20. Two-dimensional domain used in the numerical simulations. Throat height h_0 and total angle of divergence α are input parameters. The inflow channel has a total height of $4r+h_0$.

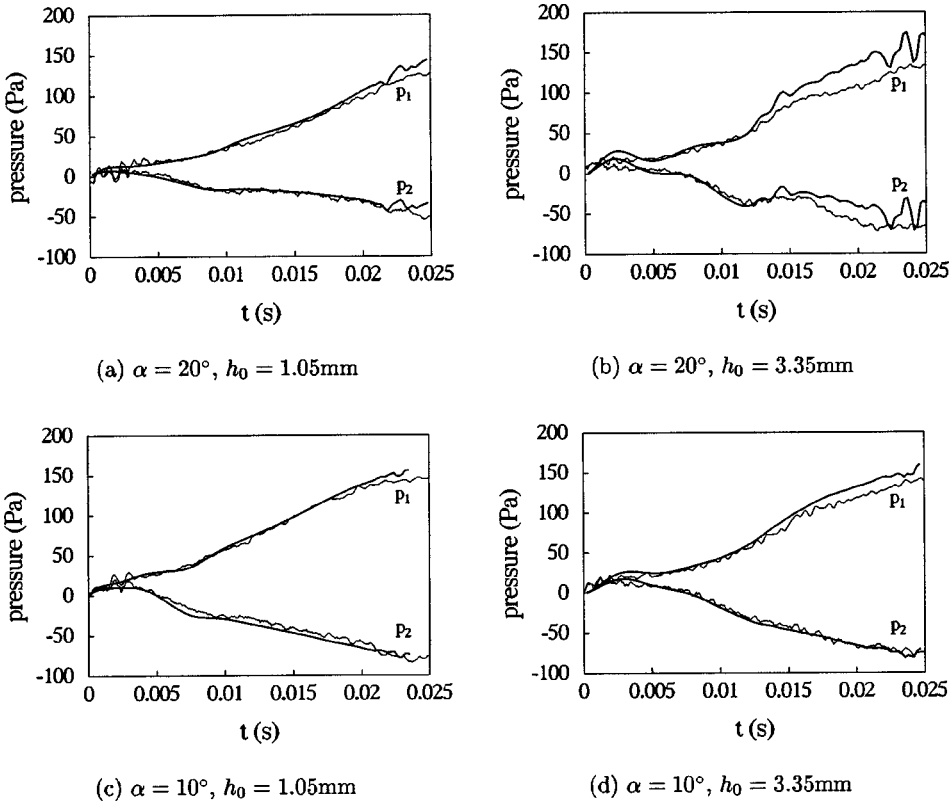


FIG. 21. Comparing numerical results for the pressures p_1 and p_2 (thick lines) to the experimental pressures (thin lines) from which the respective inflow velocities have been obtained.

ration point and of the pressure in the throat as observed in the experiments presented in Fig. 15.

After having demonstrated the complexity of this type of flow we will evaluate a simplified quasi-steady description of the flow. In Fig. 24 a close-up of the flow through the glottis is presented. The figure shows the velocity profiles in the region of the throat and illustrates the fact that a boundary-layer description is reasonable in the region around the throat. The velocity profiles exhibit a thin boundary layer near the walls and an approximately uniform profile in the bulk of the flow. Only further downstream the first large vortical structures can be observed.

Similar to Figs. 13 and 16 we present in Fig. 25 the pressure ratio $(p_2-p_s)/(p_1-p_s)$ as a function of $(p_1$

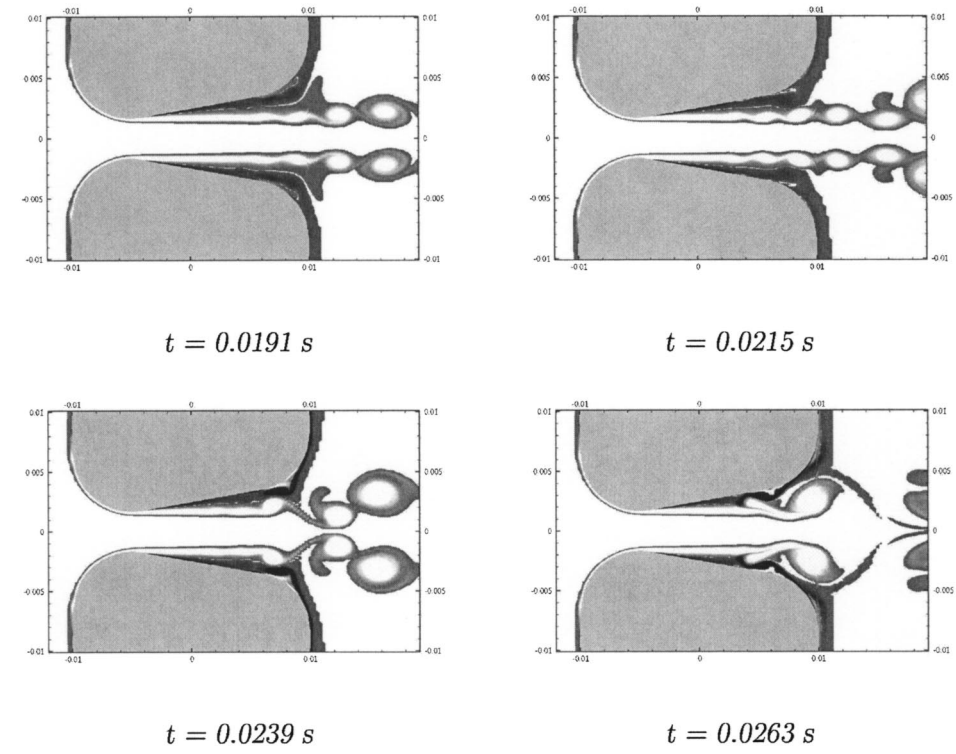
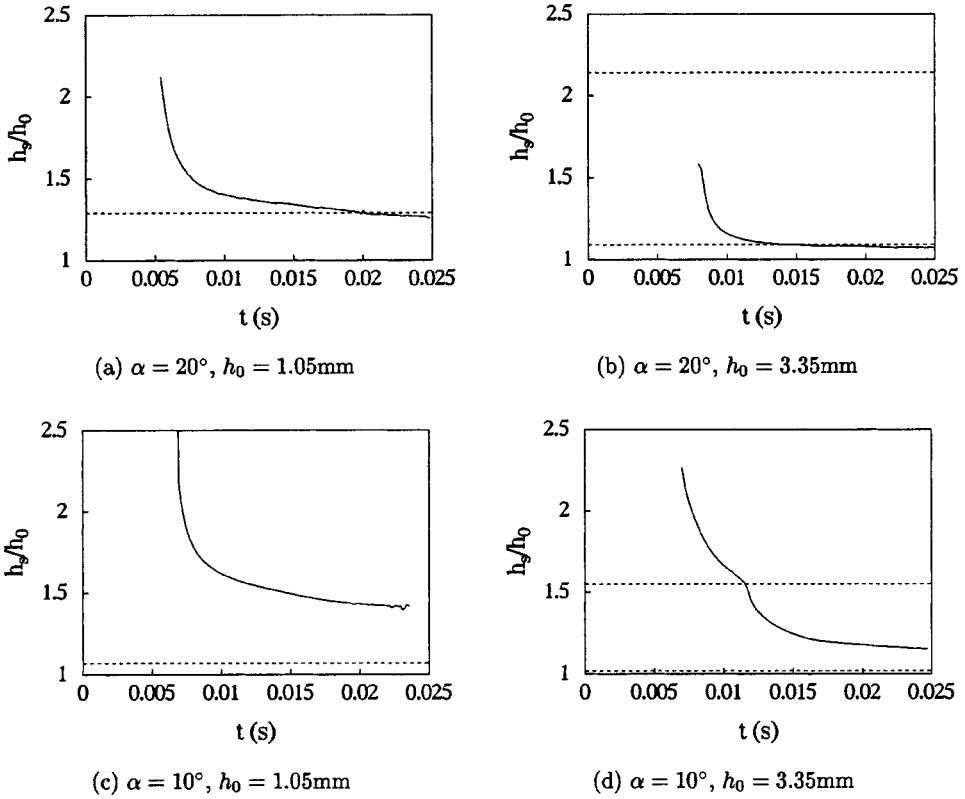


FIG. 22. Close-up of vorticity distribution in the diffuser region showing a complex interaction of vortices and the diffuser walls. Results obtained for α=20° and h_0=3.35 mm.



(a) $\alpha = 20°$, $h_0 = 1.05$mm

(b) $\alpha = 20°$, $h_0 = 3.35$mm

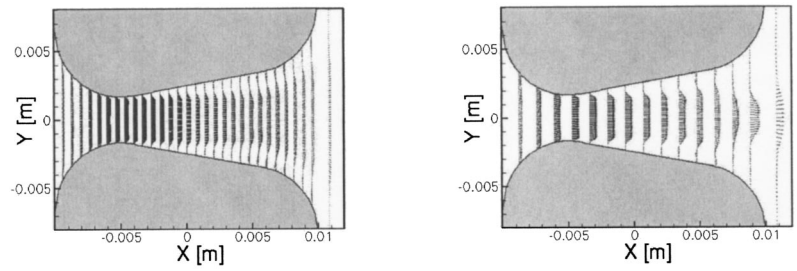
(c) $\alpha = 10°$, $h_0 = 1.05$mm

(d) $\alpha = 10°$, $h_0 = 3.35$mm

FIG. 23. Point of zero wall stress (as defined by $\omega_{\text{wall}}=0$ on the wall) as a function of time obtained from the numerical simulations. The dashed lines illustrate the starting point (lower line) and the end point (upper line) of the diffuser section. This upper line is out of the range in the left graphs for $h_0 = 1.05$ mm as it corresponds to $h_s/h_0 = 7$.

$-p_s$), remembering that p_1, p_2, and p_s can be functions of time. Although the boundary-layer model incorporates the assumption that the pressure at the separation point is equal to zero ($p_s=0$), this is not the case in the numerical simulations. The separation point in the numerical results is determined using the condition $\omega_{\text{wall}}=0$ (see Fig. 23). Then p_s is the pressure on the symmetry line at the horizontal (stream-wise) position of the separation point. Hence the ratio $(p_2-p_s)/(p_1-p_s)$ can be determined from the numerical results.

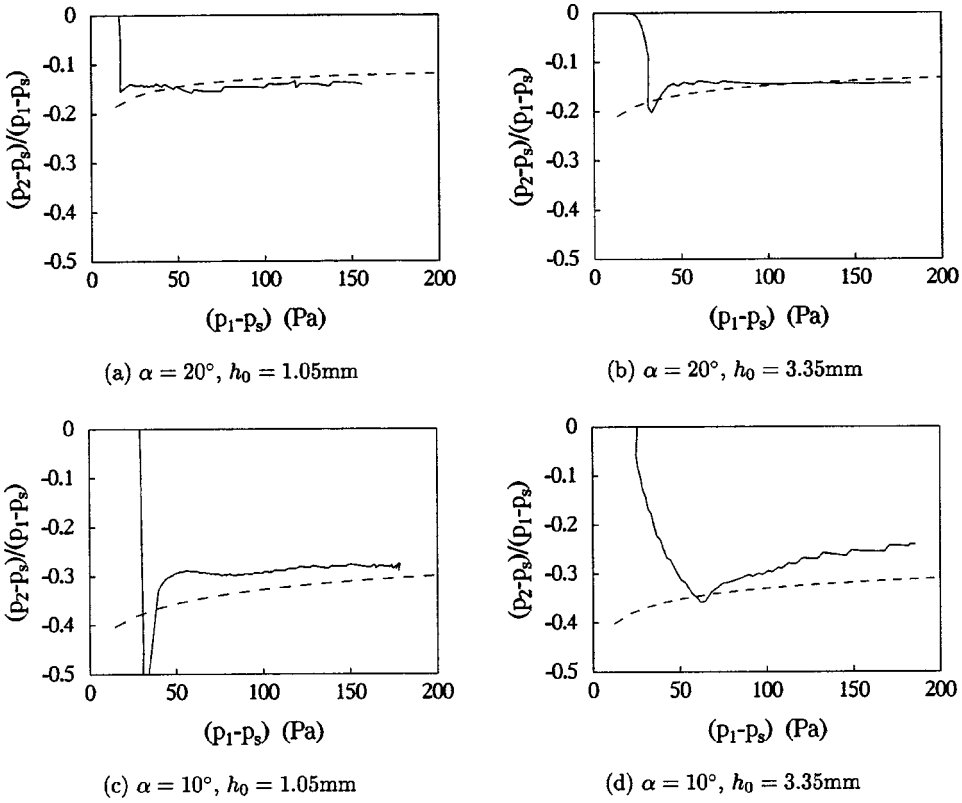
After an initial stage, in which the flow is essentially unsteady, we see that the agreement between numerical results and boundary-layer prediction is quite good. This is surprising since the boundary-layer prediction and the steady limit of the experiments did not agree so well (see Figs. 13 and 16). An explanation can be found in Fig. 26. In this figure the pressures p_1 and p_2 are plotted together with (p_1-p_s) and (p_2-p_s) in each graph. The dashed lines represent the pressure with respect to the far field, which agreed with the experimentally measured pressure as illustrated by Fig. 21. The solid smooth lines represent the pressure with respect to the separation point, as used in Fig. 25. The difference between these two lines is the contribution to the pressure due to the flow downstream of the separation point inside the constriction and to the jet flow. Note that the pressure drop p_2 from throat to separation point is about equal to the pressure drop from the separation point to the far-field. This conclusion is in agreement with the results obtained recently by Scherer *et al.* (2001) by means of steady flow numerical solution of the Navier–Stokes equations. This explains the large discrepancies found between experiments and boundary-layer theory. Note also that the pressure fluctuations are due to the flow downstream of the separation point. This indicates that the structure of the jet flow might be important to predict the source of sound in speech modeling. In particular, the jet instabilities can induce higher frequencies which are experienced in speech as broadband ''noise.''



(a) t = 11.9 ms

(b) t = 25.1 ms

FIG. 24. Close-up of the constriction region showing the velocity vectors. $\alpha=20°$ and $h_0=3.35$ mm.



(a) $\alpha = 20°$, $h_0 = 1.05$mm

(b) $\alpha = 20°$, $h_0 = 3.35$mm

(c) $\alpha = 10°$, $h_0 = 1.05$mm

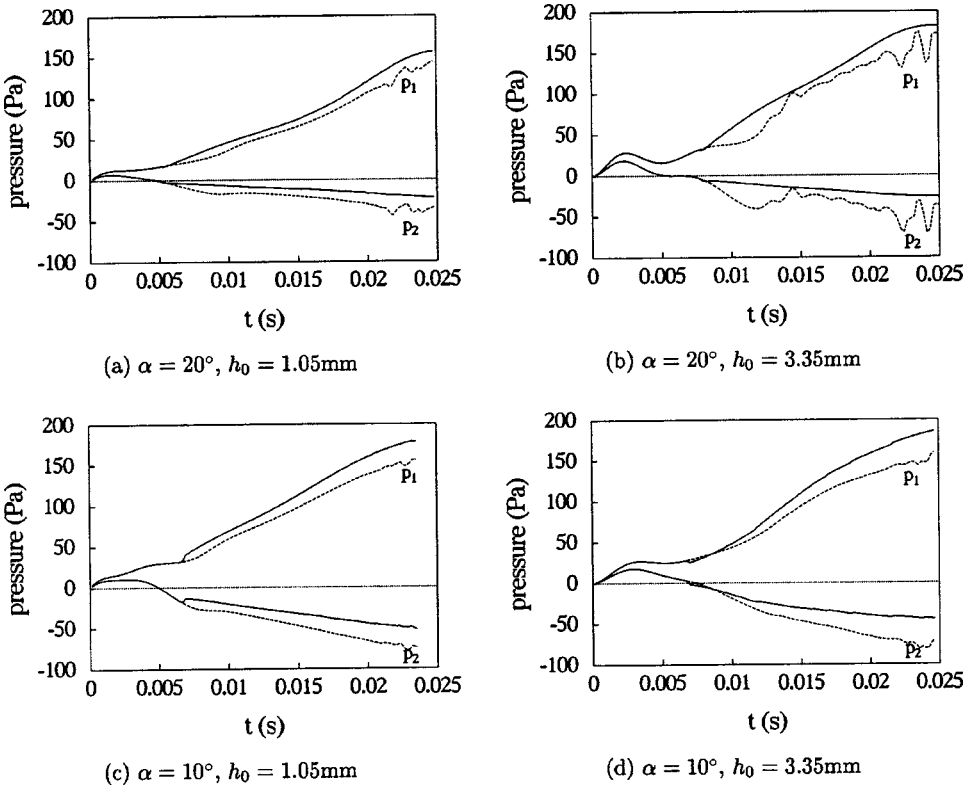
(d) $\alpha = 10°$, $h_0 = 3.35$mm

FIG. 25. Comparing numerical results (solid lines) for the pressure ratio $(p_2-p_s)/(p_1-p_s)$ to quasi-steady boundary-layer predictions (dashed lines). p_s is the pressure at the point of zero wall stress.

IV. CONCLUDING REMARKS

In this paper we set out to describe the flow through *in vitro* models of the glottis. By means of a combined experimental, numerical, and theoretical study we have managed to explain most of what was observed.

First of all, the richness of phenomena observed in the experiments demonstrates the complexity of this type of flow. However, we have shown results that make it reasonable to ignore some of these phenomena in a model of the flow through the glottis. The Coanda effect as well as the transition to a turbulent flow need in order to appear a time delay comparable to the opening or closing time of the glot-



(a) $\alpha = 20°$, $h_0 = 1.05$mm

(b) $\alpha = 20°$, $h_0 = 3.35$mm

(c) $\alpha = 10°$, $h_0 = 1.05$mm

(d) $\alpha = 10°$, $h_0 = 3.35$mm

FIG. 26. Influence of the jet on the pressure signal. Solid lines represent pressures (p_1-p_s) and (p_2-p_s) with respect to the pressure at point of zero wall stress, while the dashed lines represent pressures p_1 and p_2 with respect to the far-field pressure as used in the experiments.

tis in normal male speech production. Female speech production corresponds to typical frequencies of 200 Hz and would be even less sensitive to these effects. This conclusion is drawn, however, for rigid models of the glottis with an unsteady flow at relatively high Reynolds numbers compared to normal speech conditions.

The behavior of the diffuserlike vocal-fold models at a Reynolds number of the order 10^3 seems similar to what has been reported in literature about diffuser performance at much higher Reynolds number of the order 10^5. Similar to results reported by Kwong and Dowling (1994) obtained in a diffuser, we found experimentally and numerically that the vocal-fold model with a diffuser angle of 20° and an initial height $h_0 = 3.50$ mm exhibited very unsteady flow behavior. For the vocal-fold model with a diffuser angle of 10° and an initial height $h_0 = 3.39$ mm we found experimentally that it acts as a well-designed diffuser. This behavior is conform data on diffuser performance (Blevins, 1984).

The comparison between experiment and boundary-layer theory in combination with a quasi-steady free jet indicated that the theoretical boundary-layer model showed typical systematic errors of 30% in the throat pressure. The numerical study, on the other hand, confirmed that the boundary-layer model was not inadequate but that some of the assumptions had to be adjusted. After an initial period of essentially unsteady flow the quasi-steady laminar boundary-layer model predicts the position of the separation point with a surprising accuracy. However, the assumption of a uniform pressure in the jet is inadequate. This is confirmed by considering that the difference between the far-field pressure and pressure at the separation point is almost as large as the pressure difference between the pressure at the separation point and the pressure at the throat of the glottis. This implies that the pressure difference due to the jet is significant. This is in agreement with the steady flow calculations of Scherer *et al.* (2001). The assumption of a quasi-steady jet without pressure recovery (quasi-parallel flow implying a uniform pressure) is in fact the main source of inaccuracy in our prediction of the throat pressure. An improvement of the jet model is necessary. Such an improvement, however, is only relevant when the mechanical modelling of the vocal folds has a similar degree of sophistication (Lous *et al.*, 1998).

Numerical results obtained by means of the vortex blob method can predict the flow inside the glottis. However, downstream of the flow separation point, turbulence appears which drastically changes the character of the flow, making the numerical results useless. This effect is suppressed by flow acceleration in the attack transient but is expected to be dramatic in the deceleration phase (Hofmans, 1998).

While this paper was being revised, experiments with oscillating walls have been performed by Deverge *et al.* (2002) which globally confirm the validity of our results.

ACKNOWLEDGMENTS

Part of this research was sponsored by the Dutch Technology Foundation (Project No. ETN33.2952).

APPENDIX: POHLHAUSEN'S METHOD

Analogous to Pohlhausen (1921), we have used a third-order polynomial description of the velocity profile in the boundary layer $u(x,y)$:

$$u(x,y) = U(x)\sum_{i=0}^{3} a_i \left(\frac{y}{\delta}\right)^i,$$

in which $U(x)$ is the main flow velocity at the edge of the boundary layer, y is the coordinate perpendicular to the wall, δ is the boundary-layer thickness, and a_i are the coefficients which are functions of the coordinate x along the wall. In order to determine the coefficients the polynomial has to satisfy four boundary conditions:

$$u(x,0) = 0, \quad u(x,\delta) = U,$$

$$\left.\frac{\partial u}{\partial y}\right|_{y=\delta} = 0, \quad \nu \left.\frac{\partial^2 u}{\partial y^2}\right|_{y=0} = -U\frac{dU}{dx}.$$

Upon substitution of the polynomial in the boundary conditions the coefficients a_i are determined. Introducing the parameter $\Lambda = (\delta^2/\nu)(\mathrm{d}U/\mathrm{d}x)$ the coefficients are

$$a_0 = 0, \quad a_1 = \frac{\Lambda}{4} + \frac{3}{2}, \quad a_2 = -\frac{\Lambda}{2}, \quad a_3 = \frac{\Lambda}{4} - \frac{1}{2}.$$

Alipour, F., Fan, C., and Scherer, R. C. (**1996**). "A numerical simulation of laryngeal flow in a forced-oscillation glottal model," J. Comput. Speech Language **10**, 75–93.

Alipour, F., and Titze, I. R. (**1996**). "Combined simulation of two-dimensional airflow and vocal fold vibration," in *Vocal Fold Physiology: Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular, San Diego), pp. 17–29.

Beale, J. T., and Majda, A. (**1985**). "High order accurate vortex methods with explicit velocity kernels," J. Comput. Phys. **58**, 188–208.

Blevins, R. D. (**1984**). *Applied Fluid Dynamics Handbook* (Van Nostrand Reinhold, New York).

Cancelli, C., and Pedley, T. J. (**1985**). "A separated-flow model for collapsible tube oscillations," J. Fluid Mech. **157**, 375–404.

Chorin, A. L., and Bernard, P. S. (**1973**). "Discretization of a vortex sheet, with an example of roll-up," J. Comput. Phys. **13**, 423–429.

Chorin, A. J., Hughes, T. J. R., McCracken, M. F., and Marsden, J. E. (**1978**). "Product formulas and numerical algorithms," Commun. Pure Appl. Math. **31**, 205–256.

Deverge, M., Pelorson, X., Vilain, C., Lagrée, P.-Y., Chentouf, F., Willems, J., and Hirschberg, A. (**2002**). "Influence of the collision on the flow through in-vitro rigid models of the vocal folds," submitted for publication.

De Vries, M. (**2000**). "A voice for the voiceless," Ph.D. thesis, Rijksuniversiteit Groningen, Groningen.

Graziani, G., Ranucci, M., and Piva, R. (**1995**). "From a boundary integral formulation to a vortex method for viscous flows," Comput. Mech. **15**, 301–314.

Hirschberg, A., Pelorson, X., Hofmans, G. C. J., Van Hassel, R. R., and Wijnands, A. P. J. (**1996**). "Starting transient of the flow through an in-vitro model of the vocal folds," in *Vocal Fold Physiology: Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular, San Diego), pp. 31–46.

Hofmans, G. C. J. (**1998**). "Vortex sound in confined flows," Ph.D. thesis, Eindhoven University of Technology, Eindhoven.

Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J. **51**, 1233–1268.

Kwong, A. H. M., and Dowling, A. P. (**1994**). "Unsteady flow in diffusers," ASME J. Fluids Eng. **116**, 842–847.

Liljencrants, J. (**1993**). Private communication.

Lous, N. J. C., Hofmans, G. C. J., Veldhuis, R. N. J., and Hirschberg, A. (**1998**). ''A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design,'' Acta Acoust. **84**, 1135–1150.

Lucquin-Descreux, B. (**1987**). ''Particle approximation of the two-dimensional Navier-Stokes equations,'' Rech. Aérosp. **4**, 1–12.

Luo, X. Y., and Pedley, T. J. (**1996**). ''A numerical simulation of unsteady flow in a two-dimensional collapsible channel,'' J. Fluid Mech. **314**, 191–225.

Pedley, T. J., and Luo, X. Y. (**1998**). ''Modeling flow and oscillations in collapsible tubes,'' Theor. Comput. Fluid Dyn. **10**, 277–294.

Pedrizzetti, G. (**1996**). ''Unsteady tube flow over an expansion,'' J. Fluid Mech. **310**, 89–111.

Pelorson, X., and Hirschberg, A. (**1997**). ''*In vitro* study of the glottal and supra-glottal flow during phonation,'' in *Proceedings of the 4th Congress on Acoustics, Marseille*, Teknea, Toulouse, pp. 337–340.

Pelorson, X., Hirschberg, A., Van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y. (**1994**). ''Theoretical and experimental study of quasi-steady flow separation within the glottis during phonation. Application to a modified two-mass model,'' J. Acoust. Soc. Am. **96**, 3416–3431.

Pelorson, X., Hirschberg, A., Wijnands, A. P. J., and Bailliet, H. (**1995**). ''Description of the flow through *in-vitro* models of the glottis during phonation,'' Acta Acoust. **96**, 3416–3431.

Pelorson, X., Hofmans, G. C. J., Ranucci, M., and Bosch, R. C. M. (**1997**). ''On the fluid mechanics of bilabial plosives,'' Speech Commun. **22**, 155–172.

Pohlhausen, K. (**1921**). ''Zur Näherungsweisen Integration der Differentialgleichung der laminaren Reibungsschicht,'' Z. Angew. Math. Mech. **1**, 252–268.

Ranucci, M. (**1995**). ''Effetti del rilascio di vorticitá nell'interazione corpo-fluido,'' Ph.D. thesis, University of Rome ''La Sapienza,'' Rome.

Rosenfeld, M. (**1995**). ''A numerical study of pulsating flow behind a constriction,'' J. Fluid Mech. **301**, 203–223.

Rothenberg, M. (**1981**). ''Acoustic interaction between the glottal source and vocal tract,'' in *Proceedings of the Kurume Vocal Fold Physiology Conference*, edited by K. N. Stevens and N. Hirano (Univ. of Tokyo, Tokyo), pp. 305–328.

Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinschi, B. R., and Afjeh, A. A. (**2001**). ''Intraglottal pressure profiles for a symmetric and oblique glottis with an divergence angle of 10 degrees,'' J. Acoust. Soc. Am. **109**, 1616–1630.

Scherer, R. C., and Titze, I. R. (**1983**). ''Pressure-flow relationship in a model of the laryngeal airway with diverging glottis,'' *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. M. Bless and J. H. Abbs (College Hill, San Diego), pp. 177–193.

Tritton, D. J. (**1988**). *Physical Fluid Dynamics* (Clarendon, Oxford).

The role of contrasting temporal amplitude patterns in the perception of speech

Eric W. Healy[a] and Richard M. Warren
Department of Psychology, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201

(Received 24 September 2001; revised 27 November 2002; accepted 16 December 2002)

Despite a lack of traditional speech features, novel sentences restricted to a narrow spectral slit can retain nearly perfect intelligibility [R. M. Warren *et al.*, Percept. Psychophys. **57**, 175–182 (1995)]. The current study employed 514 listeners to elucidate the cues allowing this high intelligibility, and to examine generally the use of narrow-band temporal speech patterns. When $\frac{1}{3}$-octave sentences were processed to preserve the overall temporal pattern of amplitude fluctuation, but eliminate contrasting amplitude patterns within the band, sentence intelligibility dropped from values near 100% to values near zero (experiment 1). However, when a $\frac{1}{3}$-octave speech band was partitioned to create a contrasting pair of independently amplitude-modulated $\frac{1}{6}$-octave patterns, some intelligibility was restored (experiment 2). An additional experiment (3) showed that temporal patterns can also be integrated across wide frequency separations, or across the two ears. Despite the linguistic content of single temporal patterns, open-set intelligibility does not occur. Instead, a contrast between at least two temporal patterns is required for the comprehension of novel sentences and their component words. These contrasting patterns can reside together within a narrow range of frequencies, or they can be integrated across frequencies or ears. This view of speech perception, in which across-frequency changes in energy are seen as systematic changes in the temporal fluctuation patterns at two or more fixed loci, is more in line with the physiological encoding of complex signals. [DOI: 10.1121/1.1553464]

PACS numbers: 43.71.Es, 43.71.An [CWT]

I. INTRODUCTION

The production of speech normally involves a vibrating or noisy source which is modified by the resonances of the throat and mouth. Changes in the shape of the resonant cavities cause changes in the spectral distribution of energy, and speech sounds are usually described in these terms. Vowels, for example, are characterized by the position of the lowest three peaks in their spectrum (formant frequencies), which can span several octaves. Consonants can also be distinguished on the basis of their spectral distributions. The fricative sounds distinguishing ''see'' and ''she'' differ primarily in the spectral shape of their stochastic noiselike energy.

However, when spectral information is limited by filtering the speech signal to eliminate some frequencies while passing others, high intelligibility can remain. In a formal study of the intelligibility of speech limited to a narrow ''spectral slit'' (Warren *et al.*, 1995), it was found that narrow-band sentences exhibited very high intelligibilities, despite a lack of features traditionally thought essential for comprehension. When sentences representing everyday speech were restricted to a single $\frac{1}{3}$-octave band having steep filter slopes (96 dB/octave) and center frequencies ranging from 1100 to 2100 Hz, well over 90% of the standard scoring keywords were identified by subjects who were hearing the sentences for the first time and who had no special training. Bands in the surrounding regions down to 530 Hz and up to 4200 Hz produced scores over 60%. These results involving the high intelligibility of narrow-band sentences were confirmed and extended by Stickney and Assmann (2001).

The intelligibility function obtained in the Warren *et al.* study follows the basic pattern of band importance functions which characterize the relative importance of various frequency regions of speech. However, according to the band importance functions of the articulation index (ANSI, 1969/R1986) or speech intelligibility index (ANSI, 1997), individual $\frac{1}{3}$-octave bands in the frequency region that yielded near-perfect intelligibility in the Warren *et al.* study provide only 9% to 11% of the total information in speech. When speech is filtered to a spectral slit, energy is eliminated from the concurrent formants used to distinguish vowels. Similarly, spectral balance, formant transitions, and other spectral features traditionally thought essential for consonant recognition are missing or severely distorted. An obvious feature of the narrow-band sentence is the temporal pattern of amplitude fluctuation, which is characteristic of the particular sentence and of the center frequency of the narrow band.

Recent years have brought a greater appreciation of temporal speech information—cues provided by the pattern of amplitude fluctuation over time (for review, see Cole and Scott, 1974; Van Tasell *et al.*, 1987; Rosen, 1992). The cues present within the temporal waveform (which include envelope, periodicity/voicing, and fine-structure) are thought to be especially important for hearing-impaired listeners, who often possess relatively poor frequency resolution (cf. Moore, 1995; Tyler, 1986), but who appear to process the

[a]Author to whom correspondence should be addressed. Currently at Department of Communication Sciences and Disorders, University of South Carolina, Columbia, SC 29208. Electronic mail: ewh@sc.edu

overall temporal pattern of speech quite well (e.g. Turner *et al.*, 1995).

Much of the work examining the particular aspects of temporal fluctuations that provide linguistic information has focused on very low fluctuation rates (under approximately 20 Hz). Houtgast and Steeneken (1985) used the preservation of these low-frequency fluctuations as a predictor of intelligibility in room acoustics. Drullman *et al.* filtered the temporal fluctuations of speech, limiting them in maximum rate in one study (1994a) and in minimum rate in another (1994b). It was found that the elimination of rates below 4 Hz or above 16 Hz had little effect on speech reception thresholds when all other fluctuations were present, and concluded (1994b) that the temporal modulation spectrum can be divided into equally important parts at 8–10 Hz (but see Drullman *et al.*, 1996). Similar regions of the modulation spectrum have been considered important by Greenberg and his colleagues (Greenberg and Arai, 1998, 2001; Greenberg *et al.*, 1998; Silipo *et al.*, 1999), who proposed the ''modulation spectrogram'' as a technique to visualize the important aspects of speech unobscured by spectro-temporal details (Greenberg and Kingsbury, 1997, also see Kollmeier and Koch, 1994).

Other work has considered the role of faster fluctuations in the perception of speech. Rosen (1992) described that voicing cues are carried by temporal fluctuations of approximately 100–200 Hz, and that fine structure cues exist at rates above that. Although low fluctuation rates can be sufficient for intelligibility, studies employing speech-modulated carrier signals have demonstrated that when spectral information is limited, higher temporal rates (up to 50–200 Hz) can contribute to intelligibility (Van Tasell *et al.*, 1987; Shannon *et al.*, 1995).

Much recent interest in temporal speech information has been generated by research involving cochlear implants. Using an acoustic model of a cochlear implant, Shannon and his colleagues (Shannon *et al.*, 1995; Shannon *et al.*, 1998) reduced spectral information while retaining temporal information, not by narrow-band filtering (as in Warren *et al.*, 1995), but by reducing the spectral detail of broadband speech. The broadband spectrum was partitioned into contiguous frequency regions and the amplitude pattern of each region was used to modulate a corresponding band of noise. Because the temporal pattern of each broad region represented the average of all patterns originally present within the band, individual patterns at individual loci were lost, and only gross spectral information consisting of changes in the relative amplitudes of broad spectral regions remained. Despite these severely reduced spectral cues, high intelligibility of sentences was obtained with as few as three or four separate patterns. Dorman *et al.* (1997) confirmed and extended Shannon *et al.*'s (1995) results using both noise band and tonal carriers. They showed that sentence recognition in quiet reached an asymptote at five channels. For individual vowels, performance asymptoted at eight channels. In an earlier study, Hill *et al.* (1968) found that recognition of individual phonemes by trained listeners leveled off at roughly 75% using five to eight channels.

Due in part to an interest in hearing impairment and prostheses, temporal information in speech has often been studied in conjunction with other usually concurrent cues. A primary nonauditory cue used by impaired listeners is speechreading (lip reading), and a considerable amount of work has shown that temporal fluctuation patterns can aid this cue (e.g., Erber, 1972; Breeuwer and Plomp, 1984, 1986; Grant *et al.*, 1985; Grant *et al.*, 1991, 1994). There have also been other demonstrations that temporal speech patterns can be used in combination with other perceptual cues. Listeners hear noise shaped by the amplitude envelope of broadband words as ''more speechlike'' if the printed version of the word is concurrently available to the listener (Frost *et al.*, 1988; Frost, 1991) or when the shaped noise is presented along with the visual representation of a speaker articulating the words (Repp *et al.*, 1992). In addition to this work, Van Tasell *et al.* (1987) demonstrated some identification of speech based solely on temporal cues. Listeners could identify items from a closed-set of 19 speech syllables (known in advance by the listeners) at above-chance levels based only upon their broadband amplitude envelopes (also see Horii *et al.*, 1971; Van Tasell *et al.*, 1992; Turner *et al.*, 1995).

We know that sentences can retain intelligibility when spectral information is reduced either by filtering to a single narrow band, or by collapsing information from all frequency regions to a limited number of channels. We also know that the fluctuating amplitude pattern of speech contains linguistic information, and that listeners can use this temporal information when combined with other perceptual cues, or when tested on a small closed-set of stimuli. When broadband sentences are filtered to a spectral slit, features long thought essential for comprehension are severely distorted or eliminated. However, intelligibility can remain quite high. An obvious feature of these stimuli is the characteristic temporal pattern of amplitude fluctuation. The following experiments were designed to investigate whether this cue is sufficient to account for the observed near-perfect intelligibility of novel narrow-band sentences.

II. EXPERIMENT 1A: INTELLIGIBILITY OF SINGLE SPEECH-MODULATED TONES

In this experiment, narrow-band sentences were used to modulate the amplitude of sinusoidal tones. The resulting speech-modulated tones followed the overall fluctuating amplitude pattern of the corresponding speech band, but lacked across-frequency differences in this pattern. If the overall temporal pattern of amplitude fluctuation is responsible for the high intelligibility observed with $\frac{1}{3}$-octave sentences, then comparable scores should be obtained when listeners are presented with individual speech-modulated tones.

A. Method

1. Subjects

Thirty listeners (three groups of ten) participated in this experiment. Listeners were recruited from introductory-level psychology courses at the University of Wisconsin—Milwaukee, and received either course credit or money for participating. All were Native English speakers with no known hearing problems and were between the ages of 18 and 40 years (median age = 20 years). Considerable care was

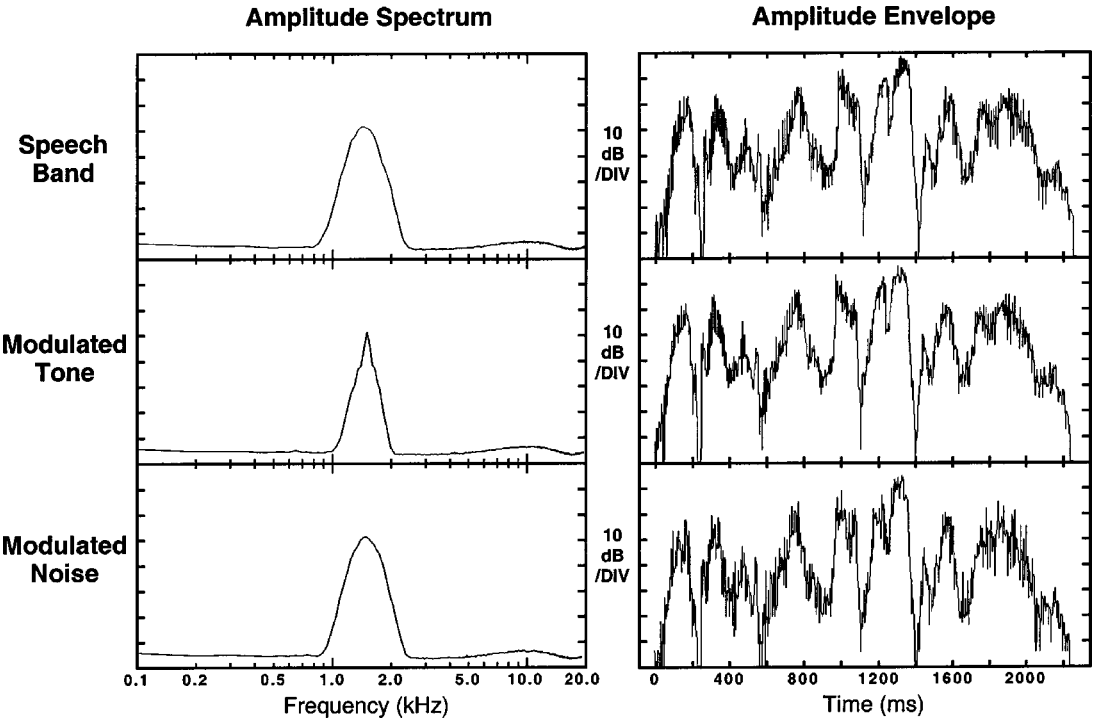


FIG. 1. Similarities between a narrow speech band and narrow speech-modulated bands: The left-most panels show the long-term average amplitude spectrum for a $\frac{1}{3}$-octave band of speech centered at 1500 Hz, a 1500-Hz tone modulated to follow the amplitude pattern of the $\frac{1}{3}$-octave speech (experiment 1a), and a $\frac{1}{3}$-octave band of amplitude-modulated noise (experiment 1b). The panels on the right show the temporal pattern of amplitude fluctuation (rms average of a sliding 2-ms window) traced by each narrow band for the example sentence, ''Walking's my favorite exercise.''

taken to ensure that no listeners had any prior exposure to the sentence materials used in these experiments. Further, participation in any one experimental condition made the subject ineligible for further participation. Thus, all listeners were hearing the stimuli for the first time. These criteria apply to all experiments described here.

2. Stimuli

The stimuli were based upon the Central Institute for the Deaf (CID) ''everyday American speech'' sentences (Silverman and Hirsh, 1955; Davis and Silverman, 1978). The 100 sentences are arranged in ten sets of ten sentences each, with 50 keywords in each set for a total of 500 keywords. An additional set of ten practice sentences was taken from the high-predictability subset of the speech perception in noise (SPIN) test (Kalikow *et al.*, 1977). The sentences were produced using natural rate and intonation by a male speaker having a General American dialect. The recordings were made from within an audiometric chamber (IAC) using a large-diaphram condensor microphone (AKG 414) and digitized at 22 kHz with 16-bit resolution using a Macintosh computer and a Digidesign analog-to-digital converter.

Speech-modulated tones were created at 750, 1500, and 3000 Hz, for comparison with $\frac{1}{3}$-octave speech bands covering a range of center frequencies which provided high intelligibility in the Warren *et al.* (1995) study. The broadband speech file was first filtered to a single $\frac{1}{3}$-octave band using a pair of cascaded eighth-order digital Butterworth filters (yielding slopes of 96 dB/octave). This filtering represents an exact digital emulation of the $\frac{1}{3}$-octave filtering accomplished using analog laboratory filters by Warren *et al.* A speech-modulated tone was created at each center frequency by multiplying a sinusoidal tone having a frequency corresponding to the center of the speech band with the full-wave rectified narrow-band speech on a sample point-by-point basis (cf. Horii *et al.*, 1971). Because the narrow-band filtering of the speech removes the higher temporal fluctuation rates, no low-pass filtering (smoothing) of the rectified speech was employed to ensure that the modulated tones were not limited in temporal rate by some arbitrary low-pass value, and were instead limited only by the narrow-band filtering employed. The modulated patterns were then refiltered to ensure that the modulation sidebands did not exceed the width of the original speech band, and that all energy lied within the original $\frac{1}{3}$-octave region. The same $\frac{1}{3}$-octave (96 dB/octave) filters used to limit the speech were employed to limit the modulated patterns in an identical manner. All processing of the stimuli was performed digitally using MATLAB and Digidesign software. The stimuli were examined at each stage of preparation using a Hewlett–Packard 3561A signal analyzer to ensure that specifications were met.

Each sentence was scaled so that the transduced level (peak of the slow-response rms average) in a flat-plate coupler was within 1 dB of 70 dBA at each ear. The stimuli were converted to analog form and presented to listeners using Digidesign digital-to-analog converters and the Macintosh computer. Analog output was routed to a Mackie (1202-VLZ) mixing board and presented diotically using matched Sennheiser (HD-250) headphones. Figure 1 shows the amplitude spectrum and an example of the temporal pattern for the speech band centered at 1500 Hz, along with those for the corresponding speech-modulated tone employed in the cur-

rent experiment and the speech-modulated noise band employed in experiment 1b.

3. Procedure

Three groups of listeners were randomly assigned to the three modulated tone conditions. Subjects were tested individually, seated across from the experimenter in an audiometric chamber. Listeners were first familiarized with the stimuli and procedures by presenting the 10 SPIN practice sentences. Each listener heard these sentences first broadband, then processed in a manner corresponding to their particular experimental condition. Each listener then heard the 100 CID sentences. They were instructed to repeat each sentence back as accurately as possible during a silent interval separating each sentence. Listeners heard each test sentence only once, received no feedback, and were encouraged to guess if unsure of the content of the sentences. Each listener heard only a single stimulus condition and participated in only a single 30-min session. The experimenter controlled the presentation of each sentence and scored the proportion of keywords reported correctly.

B. Results

A single intelligibility score based on the percentage of 500 keywords accurately reported was calculated for each listener. A group mean intelligibility score and standard error for each condition resulted from averaging these individual means. Panel (a) of Fig. 2 displays the group mean intelligibility scores and standard errors obtained for the three $\frac{1}{3}$-octave speech-modulated tones presented in the current experiment, along with data representing three of the $\frac{1}{3}$-octave speech bands presented to individual groups of subjects by Warren *et al.* (1995). The use of identical speech recordings, identical filtering parameters, the same practice and test procedures, and the same general subject pool allow the results obtained in the current experiment to be directly compared to those obtained previously.[1] In sharp contrast to the high intelligibility of narrow-band sentences, the individual speech-modulated tones exhibited scores near zero. This vast difference between scores was obtained despite the great similarity between the overall temporal fluctuation patterns of the two types of stimuli.

Listener's scores were subjected to a two-factor (2 stimulus types $\times 3$ center frequencies) analysis of variance (ANOVA) which revealed a significant main effect of stimulus type (speech band versus speech-modulated tone) [$F(1,84)=1843.5$, $p<0.0001$], and center frequency [$F(2,84)=35.6$, $p<0.0001$], and a significant interaction [$F(2,84)=38.3$, $p<0.0001$]. The critical comparison is between the speech band and the speech-modulated tone at each center frequency, and individual means comparisons indicated that these scores differed significantly at each frequency [$F(1,28)\geqslant 423.0$, $p<0.0001$]. Simple effects testing of the three narrow speech bands [$F(2,57)=75.9$, $p<0.0001$] revealed that the intelligibility score was different at each center frequency ($p<0.001$ using Scheffé's S). Although the effect of center frequency was also significant for the three modulated tone conditions [$F(2,27)=15.6$, $p<0.0001$], the scores for the 750- and 1500-Hz conditions were equivalent ($p=0.85$), but were significantly lower than that of the 3000-Hz condition ($p<0.001$), which likely accounts for the significant interaction observed in the two-factor ANOVA.

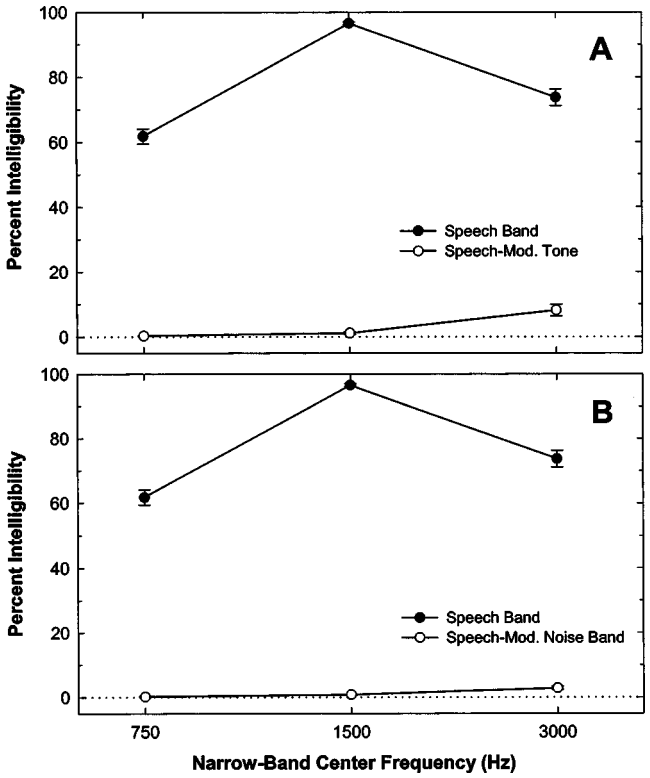


FIG. 2. The high intelligibility of sentences limited to a single $\frac{1}{3}$-octave band, versus the negligible intelligibility of corresponding speech-modulated tones (a) or noise bands (b), at each of three center frequencies: The speech-modulated signals followed faithfully the overall fluctuating amplitude pattern of the narrow speech bands, but lacked contrasting amplitude patterns within the band. Data represent group means and standard errors for 100 sentences containing 500 phonetically balanced keywords heard by each listener. Separate groups of 20 untrained listeners each heard the narrow speech bands and separate groups of ten listeners each heard the speech-modulated stimuli.

III. EXPERIMENT 1B: INTELLIGIBILITY OF SINGLE SPEECH-MODULATED NOISE BANDS

Experiment 1b was an exact replication of experiment 1a, except that speech-modulated noise bands replaced the modulated tones. Like the modulated tones, the $\frac{1}{3}$-octave noise bands followed the overall amplitude fluctuation pattern of the narrow-band speech, but lacked systematic across-frequency differences in this pattern. This replication employing a different type of stimuli and a different modulation technique was performed to ensure that the results obtained with the modulated tones was not a result of any particular manipulation performed.

A. Method

Thirty additional listeners (three groups of ten) were recruited. The speech-modulated stimuli were similar to those in experiment 1a, except that amplitude-modulated noise replaced the amplitude-modulated tone carrier signal at each center frequency (see Fig. 1). Identical speech recordings and prefiltering procedures were used to create $\frac{1}{3}$-octave speech

bands at each of the three center frequencies. The narrow-band speech was converted to speech-correlated noise by randomly reversing the polarity of individual sample points with a probability of 0.5 (Schroeder, 1968). This manipulation maintained the moment-to-moment amplitude (temporal) information within the signal, but destroyed across-frequency differences in the modulation pattern (spectral information) and produced a uniformly amplitude-modulated noise. Because the random reversal of sample points produced a broad speech-modulated noise, the same filters used to create the speech bands were employed to restrict the fluctuating noise pattern in an identical manner. Separate groups of listeners were assigned to the three center frequencies, and experienced procedures identical to those of experiment 1a.

B. Results

Like the modulated tones employed in experiment 1a, the speech-modulated noise bands exhibited near-zero intelligibility scores. Panel (b) of Fig. 2 shows the group mean intelligibility scores and standard errors for the speech-modulated noise bands presented in the current experiment, along with those for the corresponding $\frac{1}{3}$-octave speech bands [data replotted from panel (a)].

A two-factor ANOVA (2 stimulus types $\times 3$ center frequencies) revealed significant main effects of stimulus type (speech band versus speech-modulated noise band) [$F(1,84)=2060.5$, $p<0.0001$], and center frequency [$F(2,84)=37.5$, $p<0.0001$], and a significant interaction [$F(2,84)=37.1$, $p<0.0001$]. As before, the critical comparison is between the speech band and the speech-modulated noise band at each center frequency, and individual means comparisons indicated that these intelligibility scores differed significantly at each frequency [$F(1,28) \geqslant 450.0$, $p<0.0001$]. Simple effects testing revealed the same pattern of results obtained for the speech-modulated tones in experiment 1a—while the scores from the narrow speech bands differed at each of the three center frequencies, the scores from the modulated noise conditions [$F(2,27)=17.7$, $p<0.0001$] were equivalent at 750 and 1500 Hz ($p=0.46$ using Scheffé's S), but were significantly lower than the 3000-Hz condition ($p<0.001$).

A supplementary analysis was conducted to examine differences in scores resulting from the use of speech-modulated tones in experiment 1a, and speech-modulated noise bands in the current experiment. A two-factor ANOVA (2 stimulus types $\times 3$ center frequencies) revealed significant main effects of stimulus type (tone versus noise) [$F(1,54)=8.4$, $p<0.01$], and center frequency [$F(2,54)=24.6$, $p<0.0001$], and a significant interaction [$F(2,54)=6.8$, $p<0.01$]. Individual means comparisons indicated that the group mean scores for the two types of stimuli were effectively equal at 750 Hz and at 1500 Hz [$F(1,18) \leqslant 0.09$, $p \geqslant 0.76$], where all means were at or below 2%, but the score of 8% obtained at 3000 Hz in experiment 1a differed significantly from the score of 3% obtained in experiment 1b [$F(1,18)=22.0$, $p<0.0001$]. The significant interaction was likely due to this relative divergence of scores at 3000 Hz, and equivalence at the other two frequencies.

C. Discussion

The amplitude-modulation techniques employed in experiments 1a and 1b allow the elimination of spectral information from speech by creating a single fluctuating amplitude pattern that corresponds to the average of all those present within a given speech band. When this single amplitude pattern is imposed on a carrier signal, all frequency regions of the carrier follow the pattern and the signal fluctuates homogeneously. When the source signal is broadband speech, the single amplitude-modulated pattern can differ substantially from the source, due to the spread of energy from any one frequency region, to all frequency regions (in the case of modulated noise). It is this spread of energy that destroys spectral characteristics, while maintaining temporal characteristics.

Given this spread of energy and the resulting lack of spectral cues, perhaps it should not be surprising that broadband carrier signals modulated by broadband speech are not intelligible (Frost *et al.*, 1988; Frost, 1991; Shannon *et al.*, 1995), unless listeners are tested on a small closed set of items (e.g., Van Tasell *et al.*, 1987), or shown visual cues to the identity of the speech (e.g., Erber, 1972, 1979). Indeed, due to the effect of peripheral auditory filtering into an ordered series of critical bands, the single amplitude pattern averaged over broadband speech does not correspond to a pattern that would occur at any location on the basilar membrane.

Unlike broadband speech-modulated signals, it can be considered surprising that intelligibility is lost when *narrow-band* speech is converted to a narrow-band speech-modulated signal. Like amplitude-modulation, narrow-band filtering eliminates spectral information. This is accomplished not by averaging information into a single homogeneous pattern, but by eliminating energy which lays outside a narrow spectral slit. The resulting speech band fluctuates in amplitude and traces a temporal pattern which characterizes the particular passage. Similarly, a speech-modulated carrier band derived from this narrow-band speech will encompass a similar narrow band of frequencies and will fluctuate with the temporal pattern of the parent speech band.

It seemed reasonable to speculate that at least some of the high intelligibility observed with narrow-band sentences was based upon temporal information provided by the overall amplitude fluctuation pattern of the spectral slit. [Stickney and Assmann (2001) also recognized the potential contribution of the overall temporal pattern.] After all, if the speech band acts primarily as a single fluctuating pattern, then, unlike broadband signals, there should be little difference between a narrow speech band and a narrow speech-modulated carrier band. However, the current experiments demonstrate that this cue is not sufficient for intelligibility. Rather, experiments 1a and 1b suggest that even within a narrow spectral slit, speech contains spectral information that is essential for its comprehension. Viewed differently, systematic across-frequency changes in energy (spectral information) produce systematic differences in the temporal pattern of amplitude fluctuation at two fixed frequencies. Therefore, the current data suggest that contrasts between different amplitude patterns occur within a narrow speech band, and that these

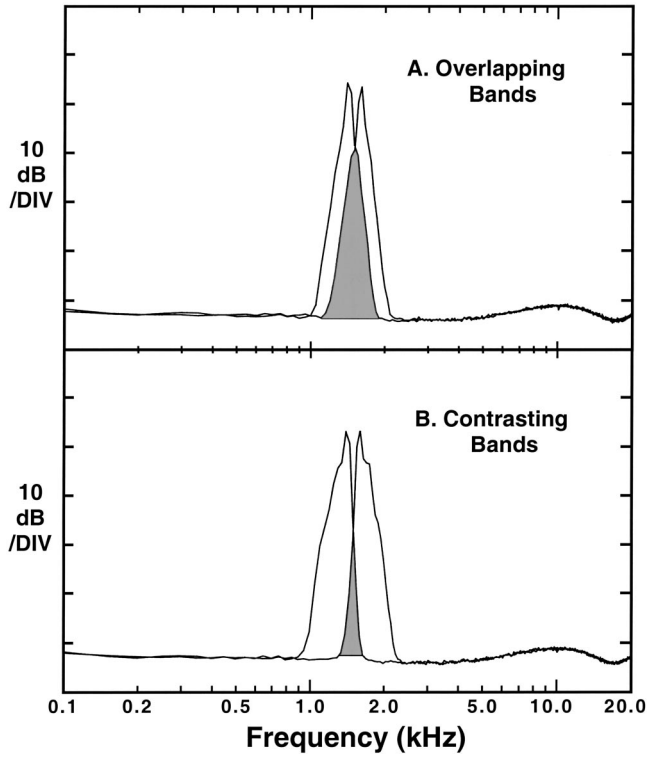


FIG. 3. Dividing a $\frac{1}{3}$-octave speech band into a pair of $\frac{1}{6}$-octave speech-modulated bands: Long-term average amplitude spectra for pairs of separately modulated $\frac{1}{6}$-octave bands (amplitude-modulated tones). The bands shown superimposed in the upper panel overlapped (shaded), disrupting each individual pattern of amplitude fluctuation. In the condition shown in the lower panel, steep filtering (480 dB/octave) was employed along the inner slopes to reduce acoustic overlap and the resultant mixing of adjacent temporal patterns.

contrasting patterns are responsible for the high intelligibility observed with narrow-band sentences.

IV. EXPERIMENT 2A: PARTITIONING A SPEECH BAND INTO A PAIR OF SPEECH-MODULATED TONES

The current experiment is a direct test of the hypothesis that the intelligibility observed with narrow-band sentences is attributable to contrasting patterns of amplitude fluctuation occurring within the narrow band. A $\frac{1}{3}$-octave band of sentences centered at 1500 Hz was divided into a contiguous pair of $\frac{1}{6}$-octaves, and these bands were used to separately modulate the amplitudes of sinusoidal tones. In one of the dual-band conditions, the $\frac{1}{6}$-octave bands overlapped spectrally, potentially disrupting each individual pattern of amplitude fluctuation [see Fig. 3(a)]. The other $\frac{1}{6}$-octave band pair employed very steep filtering to reduce acoustic overlap and to preserve each individual temporal pattern [see Fig. 3(b)]. Each $\frac{1}{6}$-octave speech-modulated band followed the overall amplitude fluctuation pattern of the corresponding speech band, but lacked within-band amplitude contrasts. However, when presented as a pair, they provide a minimal contrast.

A. Method

Forty additional listeners (two groups of 20) were recruited for this experiment. The sentences were first filtered into a pair of contiguous $\frac{1}{6}$-octave bands, corresponding to center frequencies of 1416 and 1589 Hz. The outer slopes for both pairs (the high-pass slopes of the lower bands and the low-pass slopes of the higher bands) were the same as those used to create the $\frac{1}{3}$-octave bands in experiment 1. In the overlapping condition shown in Fig. 3(a), the inner filter slopes (the low-pass slope of the lower band and the high-pass slope of the higher band) were also created using these 96 dB/octave filters and the bands intersected at 1500 Hz, overlapping considerably. In the contrasting dual-band condition shown in Fig. 3(b), the inner slopes were created using ten cascaded passes through eighth-order digital Butterworth filters, producing steep slopes of 480 dB/octave. These bands also intersected at 1500 Hz, but the cascaded filtering produced an extremely-narrow notch between the bands which served to further reduce acoustic overlap. Each $\frac{1}{6}$-octave speech band was used to modulate the amplitude of a sinusoidal tone having a frequency of either 1416 or 1589 Hz, by multiplying the full-wave rectified speech band and the tone on a sample point-by-point basis. The speech-modulated patterns were then postfiltered using the same digital filters used to create the individual $\frac{1}{6}$-octave speech bands. As before, the amplitude of each sentence in each band was adjusted so that its slow rms peak was within 1 dB of the 70 dBA presentation level.

To create the dual-band pairs, the modulated tones were mixed at equal amplitudes using the Digidesign software. To account for the effects of the frequency-dependent phase shifts of the infinite-duration impulse response (IIR) filters, and to ensure proper temporal alignment between members of each pair, the high band in the contrasting pair was delayed relative to the low band by 5 ms. No time correction was required for the overlapping pair. The dual-band pairs were converted to analog form and presented to listeners using the same apparatus employed in experiment 1. Separate groups of listeners were randomly assigned to the two dual-band conditions and heard the 100 CID sentences using procedures identical to those employed in experiment 1.

B. Results

Panel (a) of Fig. 4 shows the group mean intelligibility scores and standard errors for both dual $\frac{1}{6}$-octave band conditions, as well as those for the corresponding $\frac{1}{3}$-octave speech-modulated tone and $\frac{1}{3}$-octave speech band, both from experiment 1a. The intelligibility of the single speech-modulated tone (with its single temporal pattern) was very low, but the narrow band of speech (with its multiple contrasting temporal patterns) produced high intelligibility scores. Some improvement was obtained when the single $\frac{1}{3}$-octave pattern was divided into separately modulated $\frac{1}{6}$-octave bands which overlapped, but the acoustic mixing and resulting disruption of individual temporal patterns in the large region of overlap appears to have hindered intelligibility. However, when the individual modulation patterns of the narrow spectral slits were preserved through steep filtering, an appreciable increase in intelligibility resulted.[2] These scores were subjected to a one-factor ANOVA which revealed a significant effect [$F(3,66)=1101.2$, $p<0.0001$]. Scheffé's *post hoc* tests indicated that all four scores displayed in Fig. 4(a) differ significantly from one another ($p<0.005$).

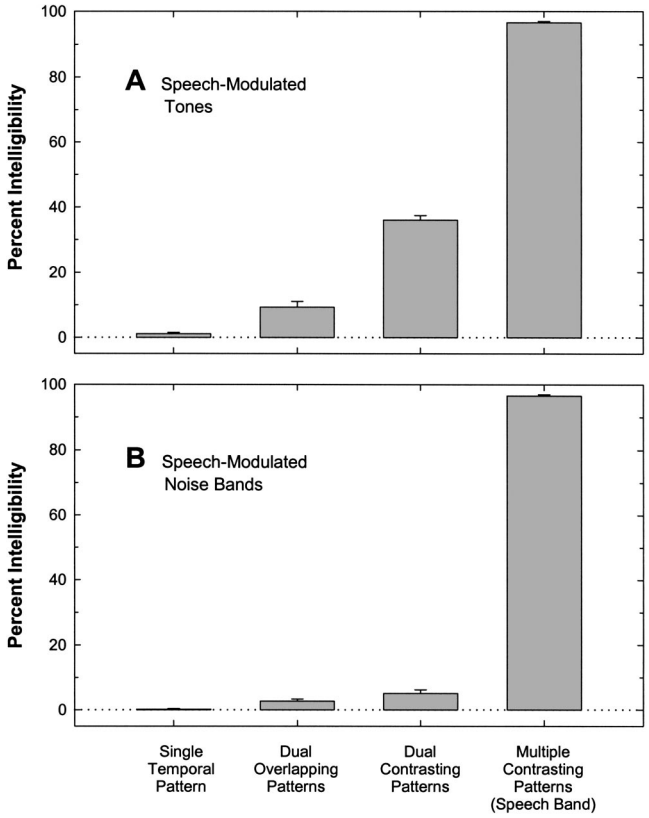


FIG. 4. Intelligibility increases resulting from increasing the amount of contrasting temporal patterns within a single narrow band: Sentences were filtered to a single $\frac{1}{3}$-octave band centered at 1500 Hz (speech band), which was used to create a single $\frac{1}{3}$-octave temporal pattern, or was divided into a pair of separately amplitude-modulated $\frac{1}{6}$-octave patterns. These $\frac{1}{6}$-octave bands either overlapped appreciably [refer to Fig. 3(a)], or had steep filtering along the inner slopes to preserve the individual contrasting patterns [refer to Fig. 3(b)]. Shown are group mean intelligibilities and standard errors for the separate groups of 20 listeners each who heard the dual-band pairs. Scores for the single temporal patterns and for the speech band are replotted from Fig. 2. Each listener heard 100 sentences containing 500 scoring keywords.

V. EXPERIMENT 2B: PARTITIONING A SPEECH BAND INTO A PAIR OF SPEECH-MODULATED NOISE BANDS

As in experiment 1, the effects observed with speech-modulated tones were confirmed using additional groups of listeners, and similar conditions in which speech-modulated noise bands replaced the modulated tones.

A. Method

Both overlapping and contrasting $\frac{1}{6}$-octave band conditions were created and presented to listeners using the same materials, procedures, and apparatus employed in the previous experiment. The same $\frac{1}{6}$-octave filtering procedures were employed to partition the narrow speech band. However, the technique of Schroeder (1968) was employed in place of the tone modulation procedure to create separately modulated noise bands that followed the overall amplitude pattern of the corresponding speech band, but which lacked within-band amplitude contrasts. Postfiltering identical to the $\frac{1}{6}$-octave prefiltering was employed to restrict each modulated noise to its frequency region of origin. Separate groups of 20 listeners each were recruited as before and randomly assigned to the two dual-band conditions.

B. Results

Panel (b) of Fig. 4 shows the group mean intelligibility scores and standard errors obtained for both pairs of $\frac{1}{6}$-octave speech-modulated noise bands, along with those for the corresponding single $\frac{1}{3}$-octave noise band and $\frac{1}{3}$-octave speech band both from experiment 1b. Although considerably lower than the scores produced by modulated tones in experiment 2a, the scores corresponding to these modulated noise bands produced the same statistically significant pattern of results: A one-factor ANOVA revealed a significant effect [$F(3,66) = 3622.0$, $p < 0.0001$], and Scheffé's *post hoc* tests indicated that all four scores differ significantly from one another ($p < 0.05$).

A supplementary analysis was performed to examine differences between the modulated tones employed in experiment 2a and the modulated noise bands employed in experiment 2b. A two-factor ANOVA (2 stimulus types $\times 2$ $\frac{1}{6}$-octave conditions) revealed significant main effects of stimulus type [$F(1,76) = 110.0$, $p < 0.0001$], and $\frac{1}{6}$-octave condition (overlap versus contrast) [$F(1,76) = 194.2$, $p < 0.0001$], and a significant interaction [$F(1,76) = 46.4$, $p < 0.0001$]. Individual means comparisons indicated that the score for the tones was significantly higher than that for the noises in both the overlapping condition [$F(1,38) = 6.8$, $p < 0.05$] and in the contrasting condition [$F(1,38) = 149.6$, $p < 0.0001$], however the difference between scores in the contrasting condition is over four times that of the overlapping condition. The difference in performance resulting from noise band and tonal carriers is perhaps due to random amplitude fluctuations of the modulated noise interfering with details of the temporal speech pattern.

C. Discussion

When narrow-band sentences were replaced with speech-modulated bands in experiment 1, near-perfect intelligibility scores fell to values near zero. Because the two types of stimuli shared similar overall temporal patterns, these results suggested that contrasting temporal patterns exist within the narrow speech bands (but not within the speech-modulated bands) and that these contrasts were responsible for intelligibility. The current experiment supports this conclusion by demonstrating that some intelligibility returned when a single narrow speech band was partitioned into a pair of separately modulated patterns. Thus it appears that contrasting temporal patterns of amplitude fluctuation residing within a narrow band of speech frequencies provide a powerful cue for recognition, and that these contrasting patterns are responsible for the high intelligibility observed with $\frac{1}{3}$-octave sentences.

VI. EXPERIMENT 3A: FREQUENCY SEPARATION OF SPEECH-MODULATED TONE PAIRS

In the following two experiments, pairs of temporal speech patterns having various frequency separations are em-

ployed to examine if the processing of contrasting temporal patterns can be extended to situations in which integration across frequencies is required. Also examined is the enhancement of across-frequency integration resulting from the addition of random noise to the spectral gap between speech-modulated bands. An additional condition was designed to examine dichotic integration of temporal speech information.

A. Method

One-hundred and twenty listeners (six groups of 20) were recruited for this experiment. Dual-band pairs were created by digitally mixing individual 70 dBA $\frac{1}{3}$-octave speech-modulated tones prepared using the procedures of experiment 1a. Five dual-band conditions were prepared, with each band pair having increasing and approximately equal logarithmic separation from 1500 Hz. The band pairs were contiguous (centered at 1336 and 1684 Hz), or were separated by (approximately) one octave (1100 and 2100 Hz), two octaves (750 and 3000 Hz), three octaves (530 and 4200 Hz), or four octaves (370 and 6000 Hz). To correct for the frequency-specific phase shifts of the IIR filters, the high band in each pair was delayed relative to the low band by a value which ranged from 0 to 25 ms and depended upon frequency separation. The delay value required to align each band pair was determined empirically by passing a single-sample click through the digital filters (combining the additive effects of pre- and postfiltering), and measuring the time between the centers of the resulting broad pulses. A dichotic condition was prepared using the modulated tone pair separated by two octaves. Rather than mixing the two bands for diotic presentation, the time-aligned signal was saved to a stereo file so that the two bands could be presented simultaneously to opposite ears. Separate groups of 20 subjects each were randomly assigned to the five dual-band spacing conditions. Each listener heard ten practice sentences and 100 CID test sentences using procedures identical to those employed earlier. The remaining 20 listeners were assigned to the dichotic condition. Ten of those heard the low-frequency band in the right ear and the high-frequency band in the left, and the other ten heard the signals to the two ears reversed.

B. Results

When the $\frac{1}{3}$-octave speech-modulated tones were presented individually in experiment 1a, single-digit intelligibility scores were obtained across all center frequencies tested. However, when the very same patterns were presented in contrasting pairs in the current experiment, appreciable scores were obtained.[3] Figure 5 shows the group mean intelligibility scores and standard errors for the 100 sentences heard by each listener in each dual-band condition. Performance was greatest when the bands were separated by one or two octave(s), and dropped off when the bands were contiguous, or were separated more widely. These scores were subjected to a one-factor ANOVA which revealed a significant effect [$F(4,95)=274.5$, $p<0.0001$]. *Post hoc* analyses using Scheffé's *S* indicated that performance was equivalent for bands separated by one and two octaves ($p=0.18$), and also for bands which were contiguous or separated by three octaves ($p=0.99$). All other comparisons were significant (p

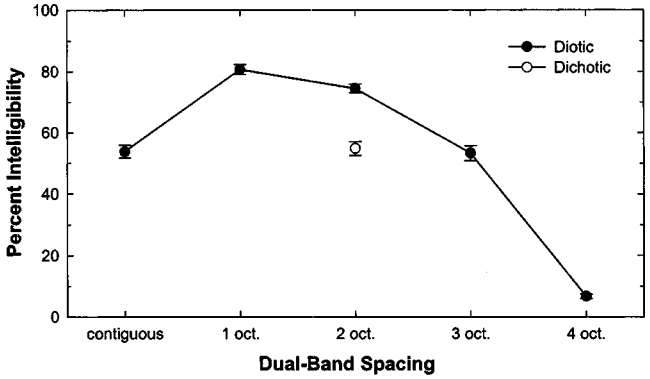


FIG. 5. Single temporal patterns, which are unintelligible when presented individually, can produce appreciable intelligibility when combined into contrasting pairs: Shown are group mean intelligibility scores and standard errors for separate groups of 20 listeners who heard pairs of temporal patterns (tones amplitude-modulated by $\frac{1}{3}$-octave sentences) separated by the values shown. Each listener heard 100 sentences containing 500 keywords. The open symbol represents performance in a dichotic condition in which an additional group of 20 listeners heard the individual temporal patterns presented to opposite ears.

<0.0001). Also displayed in Fig. 5 is the performance of the dichotic group. Although the individually unintelligible modulated tones produced appreciable intelligibility when presented to opposite ears, the score obtained under dichotic presentation (55%) was lower [$t(38)=7.6$, $p<0.0001$] than that obtained when the same patterns were presented diotically (75%).

VII. EXPERIMENT 3B: FREQUENCY SEPARATION OF SPEECH-MODULATED NOISE BAND PAIRS

This experiment was similar to experiment 3a, except that speech-modulated noise bands replaced the speech-modulated tones. In an additional set of conditions, a band of Gaussian noise was introduced to the spectral gap between speech-modulated bands to examine the increase in across-frequency integration and intelligibility associated with spectral restoration (Warren *et al.*, 1997).

A. Method

One-hundred and forty additional listeners (seven groups of 20) were recruited. One-third octave speech-modulated noise bands were created using the procedures of experiment 1b. The bands were arranged into five time-aligned pairs using the same frequency separations employed in experiment 3a. Separate groups of 20 listeners each were randomly assigned to the five dual-band pair conditions. The remaining 40 listeners were reserved for the spectral restoration conditions. They were tested using the procedures employed previously.

B. Results

Like the modulated tones, the individually unintelligible modulated noise-bands produced appreciable intelligibility when presented as contrasting pairs. Figure 6 displays the group mean intelligibility scores and standard errors based upon the 500 keywords heard by each listener, for each of the five dual-band conditions.

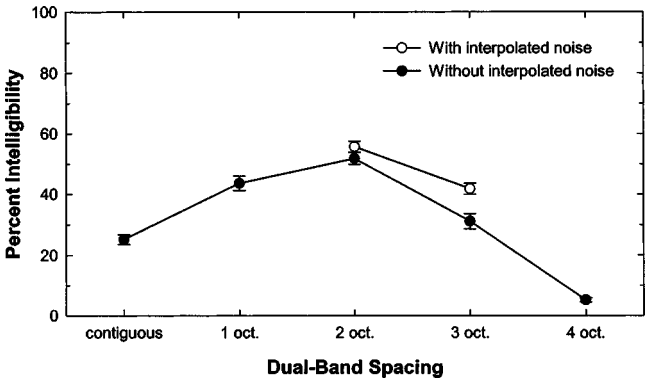


FIG. 6. As in Fig. 5, except that $\frac{1}{3}$-octave speech-modulated noise bands replaced the amplitude-modulated tones. The open symbols represent scores from additional groups of subjects who heard the speech-modulated noise bands along with an interpolated band of Gaussian noise filling the spectral gap between bands.

A one-factor ANOVA [$F(4,95)=83.6$, $p<0.0001$] and subsequent Scheffé tests revealed that, as in experiment 3a, the band pairs separated by one and two octaves produced equivalent scores ($p=0.08$), as did the pairs which were contiguous, or were separated by three octaves ($p=0.34$). All other comparisons were significant ($p<0.001$).

A supplementary analysis was performed to examine differences between the tone and noise band carrier signals. A two-factor ANOVA (2 stimulus types $\times$5 dual-band separations) revealed significant main effects of carrier signal [$F(1,190)=364.2$, $p<0.0001$], and dual band spacing [$F(4,190)=311.4$, $p<0.0001$], and a significant interaction [$F(4,190)=25.0$, $p<0.0001$]. Individual means comparisons indicated that the speech-modulated tone pairs produced higher intelligibility scores than the modulated noise bands at each dual-band separation [$F(1,38)\geqslant 71.6$, $p<0.0001$], except at four octaves where both scores were below 10% [$F(1,38)=0.3$, $p=0.56$].

C. Spectral restoration

Stochastic noise introduced to the spectral gap between speech bands that are widely separated in frequency can produce an increase in intelligibility through spectral restoration (Warren *et al.*, 1997). Like phonemic restoration, spectral restoration allows speech to be intelligible when portions of the signal are subject to masking. In the case of spectral restoration, intelligibility is restored when some speech frequencies are obliterated while others are spared. Although the addition of noise to the spectral gap provides no additional cues to the identity of speech, it is hypothesized to enhance the integration of information which is widely separated in frequency.

Experiments 1 and 2 showed that even very narrow bands of speech can contain contrasting temporal patterns, which account for their high intelligibility. In the current experiment, spectral restoration was examined using speech-modulated bands, rather than speech bands, to eliminate within-band contrasts and strictly isolate across-frequency integration. One-third octave speech-modulated noise bands separated by approximately three octaves (530 and 4200 Hz) and by two octaves (750 and 3000 Hz) were presented along with a continuous band (Wavetek Model 751A, 115 dB/octave) of Gaussian noise (Brüel & Kjær 1405) that filled the spectral gap between bands. To account for the primarily upward spread of masking, the bandwidth of the noise was adjusted so that it just masked sinusoidal tones having frequencies and amplitudes corresponding to the 30-dB down-points of the low-pass skirt of the lower bands and the high-pass skirt of the higher bands. These adjustments resulted in noise bandwidths of 830–2600 Hz for the bands separated by three octaves, and 1150–1750 Hz for the two-octave separation. The level of the noise was set to 60 dBA (rms average), the level found by Warren *et al.* (1997) to provide the maximum increase in intelligibility for speech bands presented at 70 dBA (peak of the slow-response rms average). Two groups of 20 listeners each were assigned to the two conditions. The experimental procedures were identical to those employed previously, except that the continuous interpolated noise was heard along with the speech-modulated bands.

In the previous examination of spectral restoration, noise introduced to the gap separating narrow speech bands having a fixed separation of four octaves produced an increase in intelligibility which averaged 10% across conditions. The group hearing the three-octave separation along with interpolated noise in the current experiment produced a mean intelligibility score of 42%, which is 11% above that produced by the group hearing the otherwise identical condition without noise [$t(38)=3.5$, $p<0.005$]. However, when the band separation was decreased to two octaves, listeners showed a nonsignificant intelligibility increase of 4% over their no-noise counterparts [$t(38)=1.5$, $p<0.16$] (see Fig. 6). The decrease in enhancement as frequency separation was decreased is consistent with the view that spectral restoration reflects an increased tendency to integrate widely separated information that might otherwise be treated as arising from separate sources.

D. Discussion

In experiments 1 and 2, it was shown that subjects possess the ability to effectively process contrasting temporal patterns occurring within a single narrow speech band. The current experiments indicate that these temporal patterns can also be integrated across wide frequency separations.[4] Further, experiments 3a and 3b demonstrate that random noise introduced to the spectral gap between patterns can enhance integration, and that normal-hearing listeners can effectively integrate temporal patterns presented to opposite ears.

VIII. GENERAL DISCUSSION

When broadband speech is filtered to a single narrow band, cues and features long thought essential for comprehension are severely distorted or eliminated. Knowing the rich information content of temporal speech patterns, it seemed reasonable to speculate that the narrow-band pattern of amplitude fluctuation, which remained intact following filtering, could be responsible for the high intelligibility observed with $\frac{1}{3}$-octave sentences. It was found, however, that this cue is not sufficient for intelligibility and that quite a different mechanism is involved. We report here evidence

that contrasting temporal patterns of amplitude fluctuation occurring at different frequency positions within a single narrow speech band are responsible for the high intelligibility of narrow-band sentences. When a narrow band of speech was processed to remove across-frequency spectral information, but maintain its temporal pattern of amplitude fluctuation, intelligibility fell from values near 100% to values near zero. However, when the speech band was divided into a pair of independently amplitude-modulated patterns, some intelligibility returned.

Speech is often characterized by spectral cues produced by systematic across-frequency changes in energy. However, these cues can be viewed differently: Systematic across-frequency changes in speech energy cause the temporal patterns of amplitude fluctuation to differ at different frequency positions. These patterns do not appear sufficient to encode many of the conventional acoustic speech features, especially when the patterns reside together within a narrow spectral slit. They are, however, sufficient for intelligibility of novel sentences by naive listeners, and therefore represent a powerful cue.

In experiment 2, it was found that intelligibility suffered when temporal patterns overlapped spectrally, and these results were attributed to a disruption of individual patterns. Although the mixed pattern will share attributes of its constituents, the contrasting details of the separate patterns are lost in the region of acoustic overlap. These findings are in accord with the widely held view that cochlear implant electrodes need to stimulate discrete populations of neurons to be maximally effective (cf. Wilson *et al.*, 1991). However, Shannon *et al.* (1998) reported a relatively small effect of overlap using a broadband acoustic model of a cochlear implant. To simulate electrode interaction, the authors broadened the filter slopes of the four contiguous speech-modulated noise bands. Performance remained near ceiling levels when slopes were broadened from 24 to 18 dB/octave, but fell when broadened to 6 or 3 dB/octave. Because interactions would have to be quite severe to mimic this substantial overlap, these results imply that speech recognition should be possible for implant patients even with considerable electrode interaction.

Significant differences exist between the stimuli employed by Shannon *et al.* and those employed here. While Shannon *et al.* varied only the presentation carrier bands (in order to most accurately simulate electrode interactions), in the current study, prefiltering of the speech (analysis bands) matched the postfiltering of carriers (presentation bands). However, it may also simply be that the narrow bandwidths and minimum number of channels employed in the current study, and the correspondingly lower intelligibility scores, made these stimuli more sensitive to the hindering effects of overlap. Shannon *et al.*'s standard four-channel broadband stimulus is quite robust, and scores are correspondingly high, and so it may have been better able to resist these effects.

In experiment 3, temporal patterns which were unintelligible when presented individually in experiment 1 were presented in pairs. Because each pattern lacked within-band amplitude contrasts, the contrast necessary for comprehension could only be obtained by integrating across members of the pair. Information was collected regarding the effectiveness of temporal speech information when contiguous and also when widely spaced on the basilar membrane. These conditions allowed an examination of the trade-off between (a) the greater information content of bands from the information-rich center of the spectrum, and (b) the decreased redundancy of widely spaced band pairs. According to the band importance functions of the articulation (ANSI, 1969/1986) and speech intelligibility indexes (ANSI, 1997), speech bands in the region surrounding 1500–2000 Hz provide the greatest relative contribution to intelligibility. However, when a pair of bands is employed, sampling of the speech spectrum is improved if the bands are separated in frequency. The maxima in the functions corresponding to the spacing of band pairs shown in Figs. 5 and 6 seem to reflect this trade-off. Only when bands were separated by four octaves did intelligibility fall to low values. While this may be interpreted as a limit in the ability to integrate information, it is likely that this separation pushes bands into frequency regions having relatively little effective speech information.

Psychophysical studies of envelope correlation have indicated that *same* modulation patterns can be compared across some frequency separation, but that detection of comodulation can decrease as the separation of the patterns increases (e.g., Richards, 1987, 1988; Strickland *et al.*, 1989; Moore and Emmerich, 1990; Takeuchi and Bradia, 1995; also see Richards, 1990). In contrast, the complementary temporal speech patterns employed in the current study produced maximum intelligibility when separated in frequency by one or two octaves. Unlike the randomly or sinusoidally modulated signals typically employed in studies of envelope correlation, temporal speech patterns change systematically as the frequency position of the narrow band changes. In contrast to the across-frequency *comparison* of random modulation patterns, the across-frequency *integration* of complementary temporal speech patterns appears to be quite robust.

As a demonstration of the striking synergy which occurred when individual temporal patterns were combined into contrasting pairs in the current study, the modulated tones, which together produced the highest intelligibility in experiment 3a, were presented individually. One group of ten listeners heard the speech-modulated tone at 1100 Hz, and another group heard the 2100-Hz tone, using materials and procedures identical to those employed earlier. When presented as a contrasting pair, the modulated tones produced a score of 81%. However, when presented individually, the 1100-Hz tone produced a score below 0.5% (standard error of 0.2%) and the 2100-Hz tone produced a score of 0.8% (standard error of 0.4%, see Fig. 7).

Experiments in speechreading have also demonstrated the advantage that pairs of temporal patterns can have over a single pattern. Breeuwer and Plomp (1984) showed that sentence intelligibility scores were increased from 23% for visual information alone, to an average of 49% through the addition of one temporal pattern, and to 75% with two patterns (also see Grant *et al.*, 1991, 1994). While it was reported that pilot testing indicated that the individual narrow-band envelopes were not intelligible when presented without

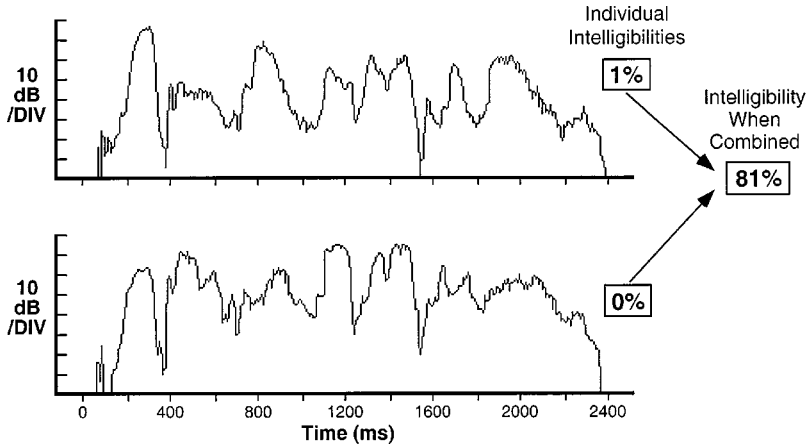


FIG. 7. Synergistic interaction of contrasting temporal speech patterns: The upper panel shows the amplitude tracing (rms average amplitude of a 10-ms sliding window) of a 2100-Hz tone amplitude modulated by a 2100-Hz $\frac{1}{3}$-octave band of speech; the lower panel shows the time-aligned amplitude tracing of an 1100-Hz tone amplitude modulated by an 1100-Hz $\frac{1}{3}$-octave band of speech, both for the sample sentence "Walking's my favorite exercise." Each intelligibility score is based upon separate groups of either 10 listeners (for each single band) or 20 listeners (for the dual-band pair). Each listener heard 100 everyday speech sentences containing 500 phonetically balanced keywords.

visual cues, pairs of these patterns produced an average intelligibility of 20% across conditions. One possible interpretation of the difference between scores obtained by Breeuwer and Plomp and those obtained in the current study involves the 20-Hz low-pass smoothing filter used in the earlier study to limit temporal fluctuation rate.

Throughout these experiments, the scores produced by the speech-modulated noise bands were lower than those obtained using speech-modulated tones. This difference may be attributed to a dilution of temporal speech information due to its mixing with the extraneous amplitude fluctuations of the noise carriers.[5] However, it is noteworthy that the temporal patterns produced by the amplitude-modulated noises were sufficiently close to those of the parent speech bands to allow appreciable intelligibility when presented as contrasting pairs in most conditions. When broad bands of speech are used to create broad amplitude-modulated noises, and when a rectification and multiplication technique is employed in place of the technique of Schroeder (1968), differences between tonal and noise carriers can be quite small (Dorman *et al.*, 1997, also see Shannon *et al.*, 1998). It is possible that the narrow bandwidths employed in the current study accentuated the random amplitude fluctuations of the noise carriers, rendering them less effective than the tones.

In addition to the ability to integrate temporal speech information across frequency separations, experiment 3a showed that listeners can effectively integrate dichotically presented temporal information. These results together indicate that peripheral interactions of temporal information is not necessary, and that these patterns can also combine centrally to produce intelligibility. The integration of information across the two ears is an essential aspect of auditory processing, and disturbances to the auditory system can affect this integration. Tests of binaural fusion (e.g., Matzker, 1959; Linden, 1964; Smith and Resnick, 1972; Willeford, 1977) have typically employed dichotic presentation of complementarily filtered bands of speech. Although the intent is to study across-ear integration, contrasting temporal patterns are present not only *across* the bands at each ear (as intended), but they are also present *within* each speech band. Because the use of speech-modulated bands (rather than speech bands) allows the strict isolation of across-band integration, they may hold promise for the examination of binaural interactions devoid of the interfering effects of contrast processing at each ear.

The current results and the conclusions drawn from them provide a novel framework for the interpretation of speech processing. It is traditional to describe acoustic speech cues according to the distribution of energy in frequency, or the position and movement of energy peaks in the frequency spectrum. Alternatively, speech information has been described according to the temporal information content of single fluctuation patterns from either narrow or broad regions of speech. However, speech is encoded by an array of neural units each responding to pressure changes occurring within a range of frequencies given by the characteristic frequency and response area of the unit. Given this encoding strategy, all information concerning complex dynamic patterns such as speech can be considered contrasting patterns. Rather than assuming that the recognition of speech is accomplished through an extraction and analysis of specific speech features in particular frequency regions, it might be worthwhile to examine correlations between various temporal patterns as speech is produced.

The results presented here provide significant challenges for existing theories of speech perception. Narrow-band speech and narrow-band speech-modulated signals share many characteristics, yet produce dramatically different intelligibility scores. Further, both of these stimuli lack traditional acoustic cues for speech perception. However, these traditional features have largely been established through an examination of individual phonemes and words, rather than speech in its normal context. Due to the robust nature of the speech signal, stimuli used in speech research must be degraded in some fashion so that intelligibility scores fall to values below the ceiling where the effect of various manipulations can be observed. This degradation of the signal has been accomplished in many ways, including filtering and by adding noise. Another traditional method of reducing speech to manageable proportions has been to study isolated phonemes and words.[6] This approach originates from the theoretical belief that the perception of speech primarily involves a "bottom-up" analysis of individual phonemes, and a subsequent synthesis into successively larger units. Although most researchers would acknowledge additional "top-down" influences provided by the context of speech, this bottom-up

construction of running speech is a basic tenet for much research.

However, speech as a means of communication normally involves the production and perception not of isolated sounds or words, but of sounds strung together into phrases, sentences, and passages that convey a meaningful message. When speech containing some of its everyday context is examined, new cues to its perception emerge. Acoustic cues necessary for the identification of isolated phonemes and words are not required for high intelligibility of everyday sentences. Further, traditional cues revealed in studies examining isolated phonemes and words cannot be considered the only cues available for recognition of speech in normal use, nor can it be safely assumed that these cues are primary in all listening conditions.

IX. SUMMARY AND CONCLUSIONS

The current experiments employed 514 normal-hearing listeners to investigate the role of contrasting temporal patterns in the perception of speech. Experiments 1 and 2 together demonstrate that speech filtered to a narrow spectral slit can retain nearly perfect intelligibility despite a lack of traditional speech features, due to contrasting temporal patterns of amplitude fluctuation within the narrow band. Experiment 3 demonstrated that the use of contrasting temporal patterns is not restricted to situations in which the patterns interact peripherally. Instead, they can be integrated across wide frequency separations, or across the two ears. It is concluded that, despite the rich information content of single temporal patterns, and despite the ability of listeners to use this information to identify speech when tested on a small closed set of items or when presented with additional cues to identity, recognition of novel sentences does not occur. Rather, intelligibility requires a contrast between at least two temporal speech patterns. These patterns can reside together within a narrow speech band, or they can be integrated across wide frequency separations or across the two ears. Speech is typically characterized by acoustic features which change in frequency as the different sounds are produced. However, a position in which across-frequency changes are viewed as changes in amplitude at fixed frequency positions is more in line with the physiological encoding of signals.

ACKNOWLEDGMENTS

These experiments were drawn from a dissertation submitted by the first author to the Graduate School at the University of Wisconsin—Milwaukee. This research was supported in part by NIH\NIDCD Grant No. DC00208 to the second author. Manuscript preparation was supported in part by NIH\NIDCD Grant No. DC01376 to Sid Bacon and by NIH\NIDCD Grant No. DC05795 to the first author. The authors thank Makio Kashino for valuable contributions throughout the course of these experiments, as well as Sid Bacon, Chris Turner, and two anonymous reviewers for thoughtful comments on a previous version of this manuscript.

[1]The high intelligibility of $\frac{1}{3}$-octave (96 dB/octave) sentences observed previously using analog laboratory filtering by Warren *et al.* (1995) was confirmed using two types of digital filters. The first was an exact digital emulation of the previously employed analog filters (dual cascaded passes through eighth-order Butterworth filters). The second was a FIR filter designed to match the amplitude spectrum of the IIR filter (82-order FIR, followed by a 750-Hz high-pass to eliminate low-frequency band-reject ripple). Both were centered at 1500 Hz and implemented in MATLAB. The mean score produced by a group of ten listeners each hearing 100 sentences in the digital IIR condition was 96.2% (standard error of 0.7%), matching the score of 96.6% (standard error of 0.4%) obtained by Warren *et al.* An additional group of ten listeners each hearing 100 sentences in the FIR condition was 98.1% (standard error of 0.4%).

[2]To confirm the results of experiments 1 and 2 under conditions that eliminate the complicating effects of filter skirts and asymmetrical filtering, narrow bands having extremely steep filter slopes were employed. A $\frac{2}{3}$-octave speech band centered at 1500 Hz having filter slopes over 1000 dB/octave (2000-order digital FIR filter) produced a mean intelligibility score of 81.6% (standard error of 1.4%) using a separate group of 20 listeners and the same 100 CID sentences. When that speech band was used to modulate a 1500-Hz tone, a score of only 1.7% was obtained (standard error of 0.4%, $n=10$). However, when the speech band was partitioned into a contiguous pair of $\frac{1}{3}$-octave bands each having extremely steep slopes, and these bands were used to modulate a pair of tones (at 1336 and 1684 Hz), an intermediate score of 31.8% was obtained (standard error of 2.2%, $n=20$) $[F(2,57)=725.8,\ p<0.0001]$.

[3]A control condition confirmed the equivalence of the current conditions to those in which a low-pass filter was employed to smooth the temporal fluctuations of the rectified speech. The dual-band pair having a separation of two octaves was recreated with the addition of 200-Hz low-pass filtering (eighth-order Butterworth) following rectification and prior to multiplication with the carrier tones. Four additional subjects heard three lists of ten sentences with low-pass present, alternated with three lists in the original low-pass-absent condition. The group mean intelligibility with low-pass present (83%) did not differ significantly $[t(3)=2.7,\ p>0.05]$ from that with low-pass absent (79%). However, when a steeper 200-Hz low-pass filter was employed, scores were reduced somewhat, indicating that temporal fluctuations contained within the filter skirt of the low-pass smoothing filter may potentially contribute to intelligibility.

[4]The 96 dB/octave filter slopes employed in experiment 3 may have caused the modulated patterns to interact at frequency positions having separations smaller than the nominal values. In order to confirm the across-frequency integration of temporal information under conditions in which interactions along the filter slopes were minimal, a pair of $\frac{1}{3}$-octave speech-modulated tones having a separation of two octaves (750 and 3000 Hz) and extremely steep filter slopes (over 1000 dB/oct) was prepared. Modulated tones were prepared as in experiment 3, except that 2000-order digital FIR filters replaced the 96 dB/octave filters in both pre- and postfiltering stages. The contrasting pair of individually unintelligible tones produced a group mean intelligibility score of 38.5% (standard error of 2.8%) using an additional group of 20 listeners, the same 100 sentences, and identical procedures, thus confirming the effective across-frequency integration of temporal information. The scores are lower in this condition when compared to the corresponding 96 dB/octave condition in experiment 3 perhaps because the nominal $\frac{1}{3}$-octave bands have different effective bandwidths due to the presence or absence of filter skirts.

[5]These temporal patterns cannot be accurately considered ''amplitude envelopes'' because they contain not only the slow fluctuations typically associated with that term, but also faster fluctuations capable of carrying (for example) voicing information. Tonal, rather than noise, carriers will better preserve these faster fluctuations.

[6]The intelligibility of narrow-band speech is reduced if isolated words or sentences having low semantic predictability are examined (Stickney and Assmann, 2001).

ANSI (**1969. R 1986**). ANSI-S3.5, 1969 (R 1986), ''American national standard methods for the calculation of the articulation index'' (American National Standards Institute, New York).

ANSI (**1997**). ANSI-S3.5, 1997, ''American national standard methods for the calculation of the speech intelligibility index'' (American National Standards Institute, New York).

Breeuwer, M., and Plomp, R. (**1984**). ''Speechreading supplemented with frequency-selective sound-pressure information,'' J. Acoust. Soc. Am. **76**, 686–691.

Breeuwer, M., and Plomp, R. (**1986**). ''Speechreading supplemented with auditorily presented speech parameters,'' J. Acoust. Soc. Am. **79**, 481–499.
Cole, R. A., and Scott, B. (**1974**). ''Toward a theory of speech perception,'' Psychol. Rev. **81**, 348–374.
Davis, H., and Silverman, S. R. (**1978**). *Hearing and Deafness*, 4th ed. (Holt, Rinehart, and Winston, New York).
Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). ''Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs,'' J. Acoust. Soc. Am. **102**, 2403–2411.
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.
Drullman, R., Festen, J. M., and Houtgast, T. (**1996**). ''Effect of temporal modulation reduction on spectral contrasts in speech,'' J. Acoust. Soc. Am. **99**, 2358–2364.
Erber, N. P. (**1972**). ''Speech-envelope cues as an acoustic aid to lipreading for profoundly deaf children,'' J. Acoust. Soc. Am. **51**, 1224–1227.
Erber, N. P. (**1979**). ''Speech perception by profoundly hearing-impaired children,'' J. Speech Hear. Disord. **44**, 255–270.
Frost, R. (**1991**). ''Phonetic recoding of print and its effect on the detection of concurrent speech in amplitude-modulated noise,'' Cognition **39**, 195–214.
Frost, R., Repp, B. H., and Katz, L. (**1988**). ''Can speech perception be influenced by simultaneous presentation of print?'' J. Memory Language **27**, 741–755.
Grant, K. W., Braida, L. D., and Renn, R. J. (**1991**). ''Single band amplitude envelope cues as an aid to speechreading,'' Q. J. Exp. Psychol. **43A**, 621–645.
Grant, K. W., Braida, L. D., and Renn, R. J. (**1994**). ''Auditory supplements to speechreading: Combining amplitude envelope cues from different spectral regions of speech,'' J. Acoust. Soc. Am. **95**, 1065–1073.
Grant, K. W., Ardell, L. H., Kuhl, P. K., and Sparks, D. W. (**1985**). ''The contribution of fundamental frequency, amplitude envelope, and voicing duration cues to speechreading in normal-hearing subjects,'' J. Acoust. Soc. Am. **77**, 671–677.
Greenberg, S., and Arai, T. (**1998**). ''Speech intelligibility is highly tolerant of cross-channel spectral asynchrony,'' Joint Meeting of the ASA and the International Congress on Acoustics, Seattle, pp. 2677–2678.
Greenberg, S., and Arai, T. (**2001**). ''The relation between speech intelligibility and the complex modulation spectrum,'' Eurospeech, Aalborg.
Greenberg, S., and Kingsbury, B. (**1997**). ''The modulation spectrogram: In pursuit of an invariant representation of speech,'' IEEE International Conference on Acoustics, Speech and Signal Processing, Munich, pp. 1647–1650.
Greenberg, S., Arai, T., and Silipo, R. (**1998**). ''Speech intelligibility derived from exceedingly sparse spectral information,'' International Conference on Spoken Language Processing, Sydney, pp. 74–77.
Hill, F. J., McRae, L. P., and McClellan, R. P. (**1968**). ''Speech recognition as a function of channel capacity in a discrete set of channels,'' J. Acoust. Soc. Am. **44**, 13–18.
Horii, Y., House, A. S., and Hughes, G. W. (**1971**). ''A masking noise with speech-envelope characteristics for studying intelligibility,'' J. Acoust. Soc. Am. **49**, 1849–1856.
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.
Kalikow, D. N., Stevens, K. N., and Elliot, 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.
Kollmeier, B., and Koch, R. (**1994**). ''Speech enhancement based on physiological and psychoacoustical models of modulation perception and binaural interaction,'' J. Acoust. Soc. Am. **95**, 1593–1602.
Linden, A. (**1964**). ''Distorted speech and binaural speech resynthesis test,'' Acta Oto-Laryngol. **58**, 32–48.
Matzker, J. (**1959**). ''Two new methods for the assessment of central auditory functions in cases of brain disease,'' Ann. Otol. Rhinol. Laryngol. **68**, 1185–1197.
Moore, B. C. J. (**1995**). *Perceptual Consequences of Cochlear Damage* (Oxford U.P., Oxford).
Moore, B. C. J., and Emmerich, D. S. (**1990**). ''Monaural envelope correlation perception, revisited: Effects of bandwidth, frequency separation, duration, and relative level of the noise bands,'' J. Acoust. Soc. Am. **87**, 2628–2633.
Repp, B. H., Frost, R., and Zsiga, E. (**1992**). ''Lexical mediation between sight and sound in speechreading,'' Q. J. Exp. Psychol. **45A**, 1–20.
Richards, V. M. (**1987**). ''Monaural envelope correlation perception,'' J. Acoust. Soc. Am. **82**, 1621–1630.
Richards, V. M. (**1988**). ''Components of monaural envelope correlation perception,'' Hear. Res. **35**, 47–58.
Richards, V. M. (**1990**). ''The role of single-channel cues in synchrony perception: The summed waveform,'' J. Acoust. Soc. Am. **88**, 786–795.
Rosen, S. (**1992**). ''Temporal information in speech: Acoustic, auditory and linguistic aspects,'' Philos. Trans. R. Soc. London **336**, 367–373.
Schroeder, M. R. (**1968**). ''Reference signal for signal quality studies,'' J. Acoust. Soc. Am. **44**, 1735–1736.
Shannon, R. V., Zeng, F. G., and Wygonski, J. (**1998**). ''Speech recognition with altered spectral distribution of envelope cues,'' J. Acoust. Soc. Am. **104**, 2467–2476.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.
Silipo, R., Greenberg, S., and Arai, T. (**1999**). ''Temporal constraints on speech intelligibility as deduced from exceedingly sparse spectral representations,'' Eurospeech, Budapest, pp. 2687–2690.
Silverman, S. R., and Hirsh, I. J. (**1955**). ''Problems related to the use of speech in clinical audiometry,'' Ann. Otol. Rhinol. Laryngol. **64**, 1234–1245.
Smith, B. B., and Resnick, D. M. (**1972**). ''An auditory test for assessing brainstem integrity: Preliminary report,'' Laryngoscope **82**, 414–424.
Stickney, G. S., and Assmann, P. F. (**2001**). ''Acoustic and linguistic factors in the perception of bandpass-filtered speech,'' J. Acoust. Soc. Am. **109**, 1157–1165.
Strickland, E. A., Viemeister, N. F., Fantini, D. A., and Garrison, M. A. (**1989**). ''Within-versus cross-channel mechanisms in detection of envelope phase disparity,'' J. Acoust. Soc. Am. **86**, 2160–2166.
Takeuchi, A. H., and Braida, L. D. (**1995**). ''Effect of frequency transposition on the discrimination of amplitude envelope patterns,'' J. Acoust. Soc. Am. **97**, 453–460.
Turner, C. W., Souza, P. E., and Forget, L. N. (**1995**). ''Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **97**, 2568–2576.
Tyler, R. S. (**1986**). ''Frequency resolution in hearing-impaired listeners,'' in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London), pp. 309–371.
Van Tasell, D. J., Greenfield, D. G., Logemann, J. J., and Nelson, D. A. (**1992**). ''Temporal cues for consonant recognition: Training, talker generalization, and use in evaluation of cochlear implants,'' J. Acoust. Soc. Am. **92**, 1247–1257.
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.
Warren, R. M., Riener, K. R., Bashford, J. A., Jr., and Brubaker, B. S. (**1995**). ''Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits,'' Percept. Psychophys. **57**, 175–182.
Warren, R. M., Hainsworth, K., Brubaker, B. S., Bashford, J. A., Jr., and Healy, E. W. (**1997**). ''Spectral restoration of speech: Intelligibility is increased by inserting noise in spectral gaps,'' Percept. Psychophys. **59**, 275–283.
Willeford, J. A. (**1977**). ''Assessing central auditory behavior in children: A test battery approach,'' in *Central Auditory Dysfunction*, edited by R. W. Keith (Grune & Stratton, New York), pp. 43–72.
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.

Quantitative evaluation of lexical status, word frequency, and neighborhood density as context effects in spoken word recognition[a)]

José R. Benkí[b)]
Department of Linguistics, University of Michigan, 4080 Frieze Building, Ann Arbor, Michigan 48109-1285

(Received 24 May 2002; revised 31 October 2002; accepted 7 November 2002)

Listeners identified a phonetically balanced set of consonant–vowel–consonant (CVC) words and nonsense syllables in noise at four signal-to-noise ratios. The identification scores for phonemes and syllables were analyzed using the *j*-factor model [Boothroyd and Nittrouer, J. Acoust. Soc. Am. **84**, 101–114 (1988)], which measures the perceptual independence of the parts of a whole. Results indicate that nonsense CVC syllables are perceived as having three independent phonemes, while words show $j=2.34$ independent units. Among the words, high-frequency words are perceived as having significantly fewer independent units than low-frequency words. Words with dense phonetic neighborhoods are perceived as having 0.5 more independent units than words with sparse neighborhoods. The neighborhood effect in these data is due almost entirely to density as determined by the initial consonant and vowel, demonstrated in analyses by subjects and items, and correlation analyses of syllable recognition with the neighborhood activation model [Luce and Pisoni, Ear Hear. **19**, 1–36 (1998)]. The *j* factors are interpreted as measuring increased efficiency of the perception of word-final consonants of words in sparse neighborhoods during spoken word recognition. [DOI: 10.1121/1.1534102]

PACS numbers: 43.71.Es, 43.71.An, 43.71.Pc [KRK]

I. INTRODUCTION AND BACKGROUND

In human speech recognition, listeners use sensory information from the speech signal to match a stimulus with an internal representation. While the structure of that internal representation, and by implication the matching process, is the subject of much debate, research has shown that the matching process is affected by many factors. These include, but are not limited to, the acoustic-phonetic properties of the stimulus, whether the stimulus is a familiar lexical item, its frequency of usage, and whether the stimulus is confusable with other potential stimuli.

A. The *j*-factor model

Boothroyd and Nittrouer (1988; also Nittrouer and Boothroyd, 1990) quantified the well-documented advantage of lexical status afforded to listeners, that items present in the mental lexicons of listeners are recognized more accurately than phonologically possible, but not meaningful, utterances. The recognition of familiar CVC words with CVC nonsense (though phonotactically English) syllables was compared using two measures of context effects that are relatively insensitive to the degree of signal degradation or overall performance level.

The present investigation uses the second of these measures, called the *j* factor, which quantifies the relationship between the recognition of a whole and the recognition of its parts. From probability theory, the recognition probability of a whole is the product of the marginal recognition probabilities of the parts. For CVC syllables, under the assumption that segments or phonemes are basic units of speech perception (e.g., Nearey, 1990; but cf. Massaro, 1998), the probability of syllable recognition p_s is the product of the recognition probabilities of the constituent phonemes

$$p_s = p_{C2} p_V p_{C2}. \tag{1}$$

Assuming the recognition probabilities p_p of individual phonemes in CVC syllables are both statistically independent and approximately equal (Fletcher, 1953), (1) can be rewritten as

$$p_s = p_p^j, \tag{2}$$

where *j* represents the number of phonemes in the syllable that are perceived independently. The *j* factor can be empirically determined by calculating the logarithms of recognition probabilities of whole syllables and segments in an identification task

$$j = \log(p_s)/\log(p_p). \tag{3}$$

A finding of $j=n$ (where $n=3$ for CVC stimuli) is consistent with the assumptions that phonemes are the basic units of speech perception and that they are independently recognized, without contextual information. The reduction of *j* below *n* is a measure of the effect of context. At the limit of $j=1$, the recognition of any one phoneme is all that is needed to recognize the whole.

For the first experiment of their original study, Boothroyd and Nittrouer prepared 120 nonsense and 120 word CVC stimuli, with both lists phonetically balanced and the targets embedded in a carrier phrase. The stimuli were

[a)]Preliminary versions of this work were presented in ''Effects of signal-independent factors in speech perception,'' Proceedings of the 28th Meeting of the Berkeley Linguistics Society, February 2002, and at a Cognitive Science Cognitive Neuroscience Workshop in August 2001 at the University of Michigan, organized by Julie Boland and Rick Lewis.

[b)]Electronic mail: benki@umich.edu

mixed with white noise at four different signal-to-noise ratios, and presented to 32 participants, blocked according to lexical status. Each participant heard and identified the 240 targets presented at one of the four signal-to-noise (S/N) ratios using a phonetic notation or standard English orthography. As expected, identification rates for word syllables were about 20%–30% higher than nonsense syllables, depending on the S/N ratio, and identification rates for phonemes in words were about 10% higher than in nonsense syllables.

The syllable and phoneme identification rates for word and nonsense targets were used to calculate *j* factors for the nonsense targets ($j = 3.07 \pm 0.14$, 95% C.I.) and the word targets ($j = 2.46 \pm 0.08$). Boothroyd and Nittrouer conclude that $j = 3.07$ for nonsense targets is consistent with perception of three independent units. The finding of $j = 2.46$ for word targets is interpreted by Boothroyd and Nittrouer as a contextual advantage such that the words used in the study were perceived as though they consisted of approximately of 2.5 independent units.

The *j*-factor reduction indicates that the higher recognition probabilities of meaningful syllables is due in part at least to the higher predictability of words relative to nonwords. As Allen (1994) says in his review of Fletcher (1953), because the set of meaningful CVC syllables is a subset of phonologically possible (e.g., nonsense and meaningful words) CVC syllables, meaningful syllables have a lower context entropy than phonologically possible syllables. Given correct recognition of part of a syllable, evaluation of ambiguous phonetic information of the rest of the syllable will be more accurate in the case of meaningful syllables because the set of possible alternatives is smaller. The *j* factor quantifies this lessening of statistical independence among the segments of meaningful syllables.

How could the reduced context entropy of meaningful words relative to nonsense words be implemented in a model of word recognition? On the basis of a computational simulation of Boothroyd and Nittrouer's experiment, Nearey (2001, in press) suggests that the *j*-factor effects could be reproduced in a Luce choice model of word recognition using a bias that favors lexical entries over nonsense words.

If the *j* factor measures bias, then manipulations of bias in a word recognition task should affect the *j* factor. It should be possible to measure *j*-factor differences for lexical biases more easily manipulated than lexical status, such as frequency of usage or response set size, as an evaluation of the claim that higher recognition rates for more frequent words result from biases (Broadbent, 1967; Norris, 1986). In the case of frequency of usage, high-frequency words are predicted to have lower *j* factors than low-frequency words.

B. Phonetic neighborhood density

In addition to frequency of usage, the existence of potential confusors to a given stimulus can affect recognition (Savin, 1963). The neighborhood activation model (NAM; Luce and Pisoni, 1998) proposes that similar-sounding words, or phonetic neighbors, compete with the actual target for activation in a Luce choice model. Degree of phonetic overlap between the neighbor and the target stimulus determines the degree of competition. Usage frequency also plays an important role, with higher values for targets increasing recognition probabilities, but higher values for the competing neighbors reducing recognition probability. These relationships are quantified in Luce and Pisoni's frequency-weighted neighborhood probability rule

$$p(ID_S) = \frac{p(S|S)\log(\text{freq}_S)}{p(S|S)\log(\text{freq}_S) + \Sigma_j p(N_j/S)\log(\text{freq}_j)}, \quad (4)$$

where $p(ID_S)$ is the probability of identifying a stimulus *S*; $p(S|S)\log(\text{freq}_S)$ is the $(\log-)$ frequency-weighted stimulus word probability of *S* given *S*, and $\Sigma_j p(N_j|S)\log(\text{freq}_j)$ is the sum of the $(\log-)$ frequency-weighted probabilities of each neighbor N_j of *S* given *S*. Luce and Pisoni's (1998) explication of the NAM states that the frequency-weighted neighborhood probability rule (4) is instantiated in individual word decision units, all of which together constitute the NAM. For the present paper, the NAM will be assumed to be represented as (4).

For empirical evaluation of the model, Luce and Pisoni use the Kucera–Francis (Kucera and Francis, 1967) usage frequencies, and their own confusion matrices of nonsense syllables in noise. The conditional probability of an item is estimated by multiplying the conditional marginal probabilities of the constituent segments obtained from the confusion matrices.

The numerator of (4), the frequency-weighted stimulus word probability, quantifies the frequency-weighted raw intelligibility of the target item. This measure is analogous to Broadbent's (1967) stimulus level α with a frequency bias term, and will be referred to as *stimulus probability* for the rest of this paper.

The denominator of (4) is the sum of the frequency-weighted stimulus probability and the frequency-weighted neighborhood probability. This latter term is a measure of the confusability of the target with other words, weighted by their log frequencies, and quantifies the *density* of the phonetic neighborhood of the target item. Targets that are highly confusable with other words have high values for this density measure, and reside in *dense* phonetic neighborhoods. Targets that are relatively not so confusable with other words reside in *sparse* phonetic neighborhoods, having low values for the density measure.

All other things being equal, accuracy of word recognition should be positively correlated with stimulus probability. On the other hand, accuracy should be negatively correlated with density, which quantifies the probability of competitors. These qualitative predictions as well as the quantitative predictions of (4) are borne out in experiments reported by Luce and Pisoni (1998).

C. Predictions

Under the assumption that *j*-factor differences are due to bias differences but not variation in the amount of stimulus information (following Nearey, 2001, in press), the *j*-factor model can be applied to the parts of (4) to evaluate whether variation in word recognition accuracy is due to a bias or due to variation in the amount of stimulus information. Simple accuracy differences reflect differences in the amount of

stimulus information, while *j*-factor differences reflect bias differences. Since the *j* factor is insensitive to the effects of stimulus information (unlike accuracy), the *j*-factor model offers a means of controlling for differences in raw intelligibility.

In the case of frequency of usage, words with high frequency of usage should have measurably lower *j* factors than words with low frequency of usage. On the other hand, words differing in stimulus probability should not have different *j* factors, all other things being equal. It is unclear what kind of an effect will be observed on the *j* factor of the product of the two, the frequency-weighted stimulus probability term in (4). It may be that the stimulus probability factor may dominate, in which case the *j* factor would not measure any context effect for stimulus probability.

On the other hand, density should be correlated with the *j* factor if the *j* factor is inversely related to bias. Consider the case of a listener perceiving partial phonetic information of a target word in a dense neighborhood. Given the partial phonetic information, the probabilities of nontarget potential responses are large, so any bias in favor of the target will be reduced. If the partial phonetic information delimits a sparse neighborhood, the probabilities of the competitors are low, and bias for the target should be high. Under this account, words in dense neighborhoods should have high *j* factors, while words in sparse neighborhoods should have low *j* factors.

The design of Boothroyd and Nittrouer offers an opportunity to test these predictions of usage frequency and neighborhood density, since the set of words was chosen to have a range of usage frequencies, and the word and nonsense syllables are phonetically balanced in the same way. While the original paper only reports average phoneme and syllable recognition probabilities for words and nonsense syllables for each participant, these probabilities could be calculated for subsets of the trials based on values of usage frequency, density, and stimulus probability. Since the nonsense syllables are phonetically balanced with the words, the segmental confusion matrices can be used to calculate conditional probabilities needed to calculate density and stimulus probability.

The effect of neighborhood density may also be considered with respect to different types of neighborhoods for a given CVC target: CV neighborhoods, which include all competitors that overlap with the first two phonemes (C_1V) of the target; VC neighborhoods, which include all competitors that overlap with the two final phonemes (VC_2); and CC neighborhoods, which include all competitors that overlap with both of the two consonants (C_1,C_2) of the target. In the computation of density described below, neighborhoods were only allowed to include competitors that overlapped with at least two phonemes of the target. With these criteria, the target *bag* would have neighbors *bad* (CV neighbor), *gag* (VC neighbor), *beg* (CC neighbor), but the words *dad* or *bed* would not be neighbors. This simplification is not unreasonable, given that failure to overlap on more than two phonemes will result in an extremely low addend in the summation of the density calculation. The simplification does allow for a straightforward means of comparing three types of phonetic overlap for CVC words.

II. METHOD

The procedure for Boothroyd and Nittrouer's experiment 1, in which participants identified CVC nonsense and word syllables at different noise levels, was followed as closely as possible, except that stimulus presentation and response collection was done online. Proportion correct of phonemes and whole syllables of different subsets of the test items were subsequently used in *j*-factor analyses.

A. Participants

Forty-three young adults were recruited from an undergraduate introductory linguistics course at the University of Michigan and participated for course credit. All were native speakers of English and reported no known hearing problems.

B. Stimuli

The same lists (in the Appendix) of CVC syllables developed by Boothroyd and Nittrouer, consisting of 120 words and 120 nonsense items, were used for this study. Both the word and nonsense syllable lists were phonetically balanced such that the phonemes in the sets of 10 initial consonants /b p d t k s h m l r/, 10 vowels /i ɪ eɪ ɛ u oʊ ɔ æ ɑ aɪ/, and 10 final consonants /p d t g k s z m n l/ were evenly distributed in the word and nonsense syllable lists.

Each item was read by the author, a native speaker of American Midwest English, in the carrier phrase ''You will write...please'' in a sound-treated room and was recorded to DAT with a Realistic Highball microphone and a Tascam DA-30 digital tape deck at a sampling rate of 48 kHz. The recording of each item embedded in the carrier phrase was converted to a WAV file at the same sampling rate and stored on computer disk. The overall level of each stimulus was adjusted so that the peak amplitudes of all stimuli were matched. This normalization did not change relative amplitudes within any given carrier phrase.

C. Procedure

The experiment was run using software running in the MATLAB (version 6.1) environment on four Windows NT laptop computers in an anechoic chamber. Signal-dependent (though uncorrelated) noise (Schroeder, 1968) was added online at one of four S/N ratios (-14 dB, -11 dB, -8 dB, -5 dB). A detailed analysis of the nonsense syllable confusion matrices reported in Benkí (to appear) shows that the nature of the confusions are highly consistent with intelligibility experiments employing white noise for both consonants (Miller and Nicely, 1955; Wang and Bilger, 1973) and vowels (Pickett, 1957; Nooteboom, 1968).

The resulting stimuli in their carrier phrases were presented for identification binaurally via AKG headphones with the volume set to a comfortable listening level, blocked according to lexical status as described below.

Thirty-seven participants were randomly assigned to one of 4 S/N ratios (11 participants at -14 dB, 9 participants at

−11 dB, 9 participants at −8 dB, and 8 participants at −5 dB). Six participants at the beginning of the study were assigned to other S/N ratios (1 participant at −10 dB, 2 participants at −9 dB, 2 participants at −7.5 dB, and 1 participant at −4 dB) in order to determine a range of S/N ratios to replicate the approximate performance levels reported in Boothroyd and Nittrouer. All 43 participants were instructed in writing that they would be listening to real and nonsense consonant–vowel–consonant syllables of English presented in a carrier phrase mixed with noise, and were to type what they heard using standard English orthography for both the words and nonsense items. A brief list of examples of English orthography for spelling nonsense items was provided.

The stimuli in their carrier phrases were presented for identification binaurally via AKG headphones with the volume set to a comfortable listening level. Following a practice block of five items, the stimuli were presented for identification, blocked according to lexical status (word or nonsense syllable). At the beginning of each experiment, the experiment software randomly determined the order of 12 word and 12 nonsense blocks. As a result, word and nonsense blocks did not necessarily alternate. Each block consisted of ten trials, which were randomly selected from the word or nonsense stimuli as appropriate. At the beginning of each block, the computer informed the participant whether the next block of ten items were either words or nonsense items, how many blocks remained, and then waited for the participant to press the ''enter'' key before presenting the first stimulus. Participants typed their responses using the keyboard, and could correct their responses for errors before pressing the enter key, but could not request the stimulus to be played again. A half-second after the enter key was pressed, the computer presented the next stimulus. Between one and four participants were run at a time, with each participant at their own computer station. Each session lasted about 30 min.

D. Phonemic analysis and scoring

The typed responses of the participants were preliminarily analyzed into phonemes with the text-to-phoneme program T2P (Lenzo, 1998). The preliminary phonemic analyses were compared to the corresponding original responses and corrected as needed.

Each phoneme response was scored as correct if it matched the corresponding stimulus phoneme, and incorrect otherwise, with the following adjustments. First, /ɑ/ and /ɔ/ were counted as matching vowels. In their stimulus list preparation and response analysis, Boothroyd and Nittrouer regarded the vowels /ɑ/ and /ɔ/ as distinct phonemes, and these distinctions were maintained in the preparation of the stimuli for the present study. However, these vowels are merged in the English spoken by many of the participants, and were therefore counted as the same vowel for scoring purposes.

Second, although the participants were told that all of the target items would consist of consonant–vowel–consonant sequences, occasionally they perceived and reported syllables containing clusters, onsetless syllables, or open syllables. The responses consisting of missing consonants in initial or final position (e.g., response ''Ed'' for target ''bed,'' or response ''bee'' for target ''beat'') were counted as ''null'' responses for the missing initial or final consonant, and the corresponding initial or final consonant scored as incorrect. Scoring was a bit more complicated for responses containing initial or final consonant clusters, with an effort made to count as much of the response as correct if appropriate. If neither consonant in the initial cluster matched the initial consonant of the stimulus, then the response was counted as an ''other'' response for the initial consonant and scored as incorrect. If the response contained an initial cluster whose second member matched the stimulus initial consonant (e.g., response ''stack'' for target ''tack''), the response cluster corresponding to the initial consonant was counted as ''other'' and scored as incorrect. If the response contained an initial cluster whose first member matched the stimulus initial consonant but the response vowel was incorrect (e.g., response ''black'' for target ''beak''), then the initial consonant was scored as correct, and the vowel response was counted as ''other'' and scored as incorrect. For responses containing a matching initial consonant, an epenthetic consonant, and a matching vowel (e.g., response ''black'' for target ''back''); half were counted as having an ''other'' response for the initial consonant but a matching vowel, and half were counted as having a matching initial consonant but an ''other'' response for the vowel. The same procedure was followed for final clusters.

III. RESULTS

A. Syllable and phoneme recognition rates

Using the above criteria, the observed probabilities of correct recognition of nonsense phonemes, word phonemes, nonsense syllables, and word syllables for each participant were calculated. The probabilities as a function of S/N ratio, averaged over the participants who performed the task at the same S/N ratio, are plotted in Fig. 1 with error bars indicating ±1 standard deviation. Although the data from all participants are in the *j*-factor analyses below, only the data from the 37 participants who performed the experiment at the four selected S/N ratios are represented in Fig. 1.

As expected, proportion correct of phonemes and syllables is higher for the word condition than that of nonsense condition at each S/N ratio. The range and pattern of performance is quite similar to those reported by Boothroyd and Nittrouer. The S/N ratios are about 11 dB lower in the present study, which may be due to differences in the quality of the stimuli, type of noise (they used spectrally shaped white noise that was the same level for all of the stimuli of a given S/N ratio, instead of the signal-correlated noise used here), or the experimental procedure.

B. Analyses of *j* factors by subjects

The values summarized by Fig. 1 were used to calculate the *j* factors for words and nonwords. For the frequency, density, stimulus probability, and frequency-weighted stimulus probability measures, the words were divided into high and low (dense and sparse for the density conditions) subsets according to whether the particular value of each word fell

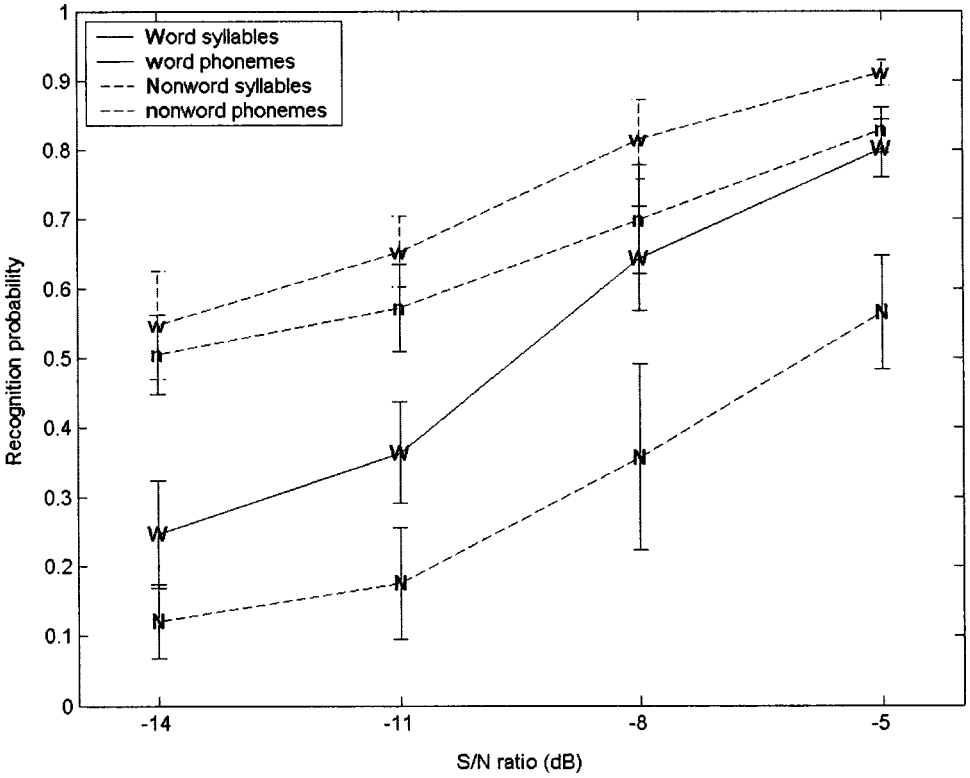


FIG. 1. Average phoneme recognition probabilities p_p in words (w) and nonwords (n), and average syllable recognition probabilities p_s for words (W) and nonwords (N). Averages are computed over participants at the S/N ratios of −14, −11, −8, and −5 dB, with error bars representing ±1 standard deviation.

above the median value, or below or equal to the median value. Frequency of usage was taken from Kucera and Francis (1967). For the most part, the high and low subsets of each measure were not unduly phonemically unbalanced, with at least one each of C_1, V, and C_2 phoneme in the high and low subsets. Exceptions were the high VC density subset with no instances of /g/ for C_2 (but 11 instances of /d/), and the low stimulus probability and frequency-weighted stimulus probability subsets with no instances of /m/ or /n/ for C_2. This latter asymmetry is not unexpected given that place of articulation in syllable-final nasals of the nonword syllables is a particularly difficult contrast for listeners to perceive (Hura *et al.*, 1992; Beddor and Evans-Romaine, 1995; Wright, 2001).

An online version of *Webster's Pocket Dictionary* (*Webster's Seventh Collegiate Dictionary*, 1967; Nusbaum, Pisoni, and Davis, 1984) was used to determine the neighbors for each target word. As discussed above, all neighbors differed with the target by one segment, with a substitution or a deletion for the third nonmatching segment, but no insertions. Thus, for the target "rice" (/raɪs/), the neighborhood would include "lice" (/laɪs/, substitution of C_1), "race" (/reɪs/, substitution of V), "ride" (/raɪd/, substitution of C_2), "rye" (/raɪ/, deletion of C_2), and "ice" (/aɪs/, deletion of C_1), but not "writes" (/raɪts/, insertion of /t/). Confusion matrices were calculated for C_1, V, and C_2 by collapsing the nonsense syllable responses across all participants. The cells of each confusion matrix were used to calculate the conditional probabilities needed to compute density and stimulus probability for each target word, using log Kucera–Francis word-frequency values as frequency weights for both targets and their neighbors.

Proportions correct of phonemes and syllables for the word trials, nonword trials, and for the high and low conditions of each measure applied to the word trials were converted into *j* factors for each participant using (3), and are plotted in the different panels of Fig. 2. The resulting *j* factors were then averaged across participants for estimates of *j* for each condition, which are also plotted in the panels of Fig. 2. Because measurement errors for probabilities near zero or unity have a large effect on the estimate of the *j* factor, if either the phoneme or syllable probability was less than 0.05 or greater than 0.95, the resulting *j* factor was not included in the calculation of average *j* factor or subsequent statistical tests, following Boothroyd and Nittrouer. The values from two participants were excluded from the *j*-factor analysis of nonwords because the syllable probabilities were too low.

1. Lexical status

The word and nonword points and curves representing the average word and nonword *j* factors are displayed in the same panel as a measure of the effect of lexical status on the *j* factor. Consistent with Boothroyd and Nittrouer's findings, $j=3.07$ with a 95% confidence interval of ±0.08 (the C.I. here and in the rest of the paper were obtained by multiplying the standard error by the upper 5% cutoff value of Student's *t* with $df=N-1$ subjects) does not significantly differ from $n=3$ [$t(40)=1.69$, $p=0.10$], as predicted by independent perception of the phonemes of each syllable. A comparison of the high and low conditions of each measure is shown in Table I, using a paired t-test to assess significance of the difference, with $df=N-1$ subjects who provided reliable estimates of the *j* factor for both conditions. Words and nonsense syllable *j* factors, are compared in the first row of Table

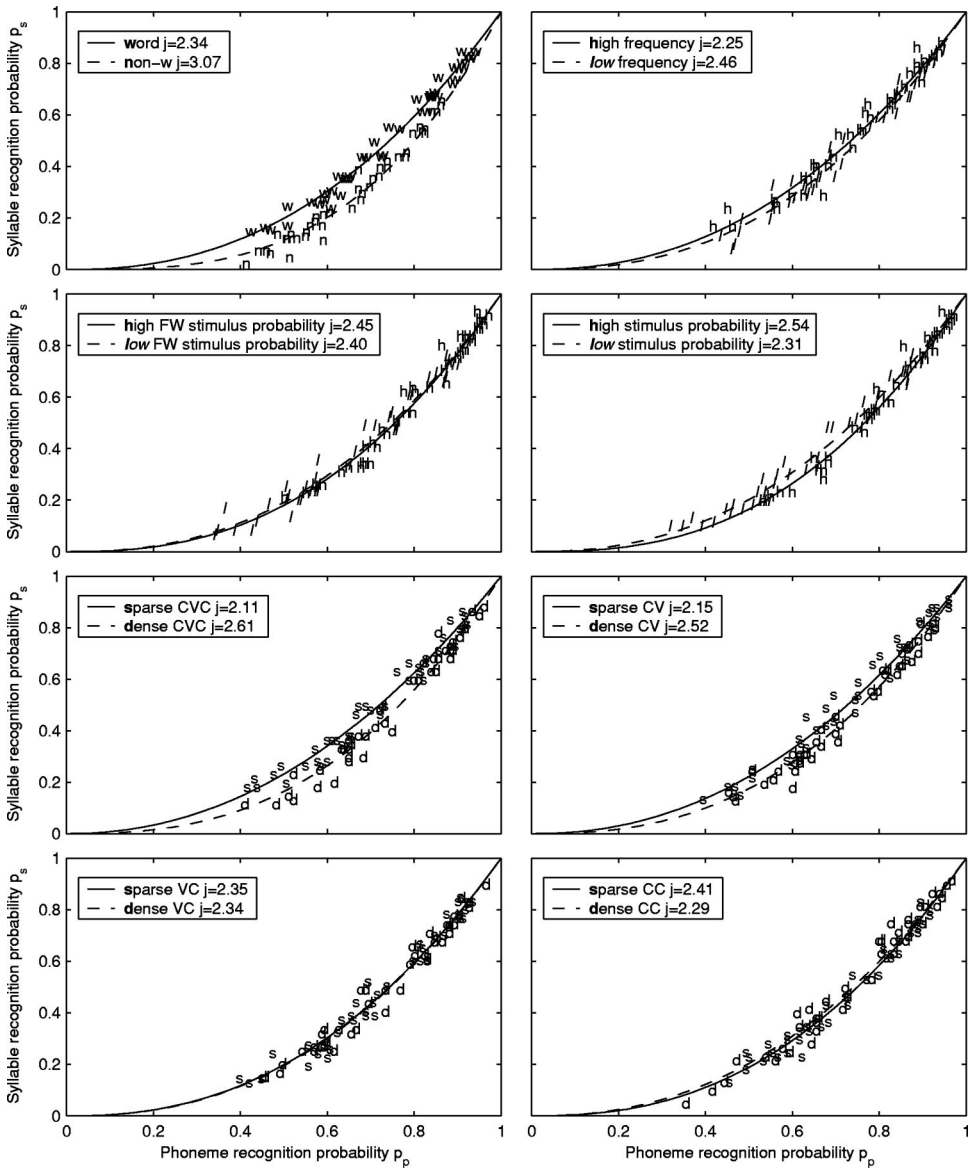


FIG. 2. Syllable recognition probability p_s as a function of phoneme recognition probability p_p for individuals. Each panel displays a pair of points for each individual, representing the probabilities for that individual averaged over a specific set of stimuli as indicated in each panel. Word (w) and nonword (n) probability points are plotted in the upper left panel, while each of the other panels display points corresponding to high (h) and low (*l*) subsets of the words for log-frequency, stimulus probability, frequency-weighted stimulus probability, overall density, CV density, VC density, and CC density [dense (d) denoting high density, and sparse (s) denoting low density]. The mean *j* factors, averaged over the points meeting the cutoff criteria for each set of stimuli, are plotted as $p_s = p_p^j$.

I, and the word *j*-factor mean, $j = 2.34 \pm 0.08$, is significantly less than the nonsense syllable *j*-factor mean. The tests in Table I are all *a priori* motivated and significance is declared using a criterion of $\alpha = 0.05$. That the word *j* factor is substantially less than the number of segments indicates that the phonemes in words are perceived not quite independently of each other, or that listeners are biased to perceive lexical items. While the *j* factor for words is lower than that found by Boothroyd and Nittrouer, the confidence intervals overlap.

Individual *j* factors are well predicted by using average *j* factors in (2) to predict syllable recognition probability p_s from phoneme recognition probability p_p for both words and nonwords. The root-mean-squared error (RMSE) for words is 0.030, and for nonwords is 0.027. Also consistent with Boothroyd and Nittrouer, individual *j* factors are not significantly correlated with phoneme recognition probability p_p for either nonsense syllables [$r^2 = 0.065$, $F(1,40) = 2.71$, $p = 0.117$] or for words [$r^2 < 0.001$, $F(1,42) < 0.001$, $p < 0.999$]. The lack of correlation with phoneme recognition probability and the good fit across the range of recognition probability supports the use of the *j* factor as an index of context effects independent of phoneme recognition probability.

2. Word frequency

Word frequency effects were measured by dividing word trials into high- and low-frequency groups using the median log Kucera–Francis frequency of all the words (3.29) as a cutoff. The high-frequency words have a mean log Kucera–Francis frequency of 4.90, while the low-frequency words have a mean log Kucera–Francis frequency of 2.46. Average phoneme and syllable recognition probabilities were calculated for high- and low-frequency words for each participant, and converted to *j* factors. As expected, the high-frequency words have a statistically significant ($p < 0.001$) lower *j* factor ($j = 2.25$) than the low-frequency words ($j = 2.46$).

3. Neighborhood density

As with frequency, median values of neighborhood density were used as cutoffs to divide the target words into dense (high density) and sparse (low density) conditions. The

TABLE I. Evaluation of the *j*-factor differences for each effect. Each row represents a comparison of the high and low (or dense and sparse) conditions of a particular measure, evaluated by a paired-comparison t-test of the *j* factors in the different conditions. The mean difference (averaged across participants) is shown with a 95% confidence interval, the standard error of the difference, and a significance value for whether the difference between the two conditions is significant.

Comparison		Mean difference (95% C.I.)	S.E.	df	*p*
Lexical status	*j*(nonwords)-*j*(words)	0.732 (0.600, 0.865)	0.065	40	<0.001
Frequency	*j*(low freq. words)-*j*(high freq. words)	0.217 (0.102, 0.332)	0.057	42	<0.001
Stimulus probability	*j*(low probability)-*j*(high probability)	−0.201 (−0.345, −0.056)	0.071	36	0.008
FW stimulus probability	*j*(low FW probability)-*j*(high FW probability)	0.078 (−0.072, 0.229)	0.074	37	0.300
Density	*j*(dense words)-*j*(sparse words)	0.507 (0.389, 0.625)	0.058	41	<0.001
CV density	*j*(dense CV words)-*j*(sparse CV words)	0.376 (0.259, 0.493)	0.058	39	<0.001
VC density	*j*(dense VC words)-*j*(sparse VC words)	0.001 (−0.096, 0.096)	0.048	41	0.996
CC density	*j*(dense CC words)-*j*(sparse CC words)	−0.122 (−0.224, −0.018)	0.051	39	0.022

words were similarly divided into dense and sparse conditions for CV, VC, and CC density measures. Figure 3 shows histograms of the distribution of each density measure. Each word had an average of 20.8 (s.d.=4.8) neighbors, broken down as 7.1 (s.d.=2.4) CV neighbors, 5.7 (s.d.=2.2) VC neighbors, and 8.0 (s.d.=3.5) CC neighbors. Individual *j* factors and average values are plotted in Fig. 2 for overall density as well as the subdivided density measures. Table I shows comparisons of the dense and sparse conditions of each density measure.

The difference in *j* factors between the dense and sparse conditions for overall density, 0.507, is significant ($p<0.001$), as well as the difference for CV density, 0.376 ($p<0.001$). As expected, according to the notion that words in sparse neighborhoods will have a favoring bias, the sparse conditions have lower *j* factors than the dense conditions.

However, the difference between the dense and sparse conditions for CC density, 0.122, is slight and in the opposite direction as anticipated, but significant ($p=0.022$). No statistically significant difference was found for the VC density measure ($p=0.996$).

4. Stimulus probability

As with frequency and density, the word trials were divided into high and low conditions for stimulus probability and frequency-weighted stimulus probability by using the median values as cutoffs. Individual *j* factors and curves of

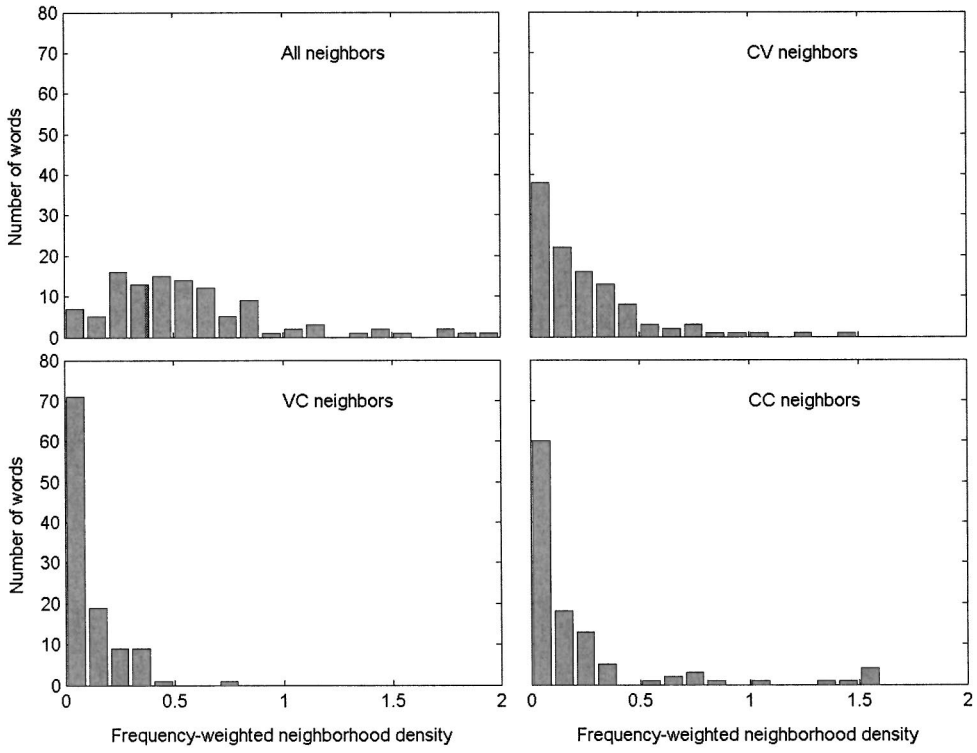


FIG. 3. Histograms of frequency-weighted neighborhood density values for the 112 words. The upper left panel shows the distribution of density calculated over all neighbors, while the other panels shows the distribution of density calculated over a specific subset of all neighbors, namely CV neighbors, VC neighbors, and CC neighbors.

TABLE II. Comparison of the magnitudes of the *j*-factor effects. Each row represents a comparison between two *j*-factor differences reported in Table I, evaluated by a paired-comparison t-test. The mean difference (averaged across participants) is shown with a 95% confidence interval, the standard error of the difference, and a significance value for whether the effect magnitudes are significantly different.

Comparison	Mean difference (95% C.I.)	S.E.	df	*p*
Lexicality—frequency	0.544 (0.376, 0.713)	0.083	40	<0.001
Lexicality—stim. prob.	0.571 (0.331, 0.812)	0.118	34	<0.001
Frequency—stim prob.	−0.007 (−0.212, 0.199)	0.101	36	0.948
Lexicality—density	0.245 (0.062, 0.428)	0.090	39	0.010
Density—stim. prob	0.306 (0.086, 0.525)	0.108	35	0.008
Density—frequency	0.291 (0.137, 0.445)	0.076	41	<0.001
Density—CV density	0.124 (−0.027, 0.275)	0.075	39	0.105
Density—CC density	0.386 (0.198, 0.575)	0.093	39	<0.001
CV density—CC density	0.259 (0.136, 0.382)	0.093	38	<0.001

the average *j* factors for the high and low conditions of both measures are plotted in the lower panels of Fig. 2. The effect of stimulus probability is significant, with the low stimulus probability condition having a *j* factor lower by 0.201 than the high condition ($p=0.008$). However, the difference between high and low conditions for frequency-weighted stimulus probability, 0.078, is not significant ($p=0.300$).

5. Comparisons of effect sizes

The *j*-factor effect magnitudes for the different measures can also be compared. Table II shows the results of selected paired-t tests of the different measures. Each row shows the mean of the difference across subjects and the *p* value for assessing whether the mean is reliably different than zero. For example, the first row compares lexicality with frequency by testing [*j*(nonwords)-*j*(words)]-[*j*(low freq. words)-*j*(high freq. words)]. For comparisons involving stimulus probability and CC density, the magnitude of these two effects was defined as *j*(high probability)-*j*(low probability) and *j*(sparse CC words)-*j*(dense CC words), respectively, since the direction of these effects was the reverse of the others.

The paired comparison t-tests reported in Table II on the significance of the difference of differences of the *j* factors in Table I indicate that the effect of lexicality is stronger than any of the other effects, 0.544 higher than frequency, 0.571 higher than stimulus probability, and 0.245 higher than density. The effects of frequency and stimulus probability are no different in magnitude ($p=0.948$), presumably canceling each other out in the frequency-weighted stimulus probability measure. The effect of density is in turn 0.291 higher than the effect of frequency and 0.306 higher than stimulus probability ($p=0.008$). Among the density measures, overall density is no different than CV density ($p=0.105$), but both overall density and CV density are stronger than the effect of CC density (0.386 and 0.259, respectively).

More conservative criteria for statistical significance might be justified for the comparisons of Table II than those of Table I, particularly since there are nine comparisons. If the tests between the different density measures are grouped into one family of three tests and the other tests into a family of six tests, Bonferroni-adjusted significance criteria would be $\alpha=0.05/3=0.0167$ and $\alpha=0.05/6=0.0083$, respectively. These adjustments do not change any of the declarations of significance.

C. Analyses of *j* factors by items

The subject *j*-factor analyses using median splits indicate that lexical status, frequency, stimulus probability, and neighborhood density—particularly CV density—influence word recognition as biasing context effects independent of phoneme recognition probability. This effect was not found for frequency-weighted stimulus probability or VC neighborhood density.

However, the histograms in Fig. 3 call into question whether a comparison of high and low (or dense and sparse) conditions using median splits is a reliable test of the effect for all of the density measures. As shown in Fig. 3, while overall neighborhood density has a fairly even distribution, the distributions of the CV, VC, and CC density measures have long right tails. This skewing is most pronounced for the distribution of VC density, due to the lower intelligibility of C_2 relative to C_1 in nonsense syllables. Because recognition probability of C_2 is lower than C_1 in the nonword confusion matrices, the distribution of VC density is further to the left of the distribution of CV density.

While the VC density measure had no effect on the *j* factor in the subjects analysis, a more rigorous evaluation would be to test for correlation between the density values of individual words and the word *j* factors averaged across subjects. Despite a skewed distribution, there should be enough variation in the density measures to indicate whether they are correlated with the *j* factor at all. This same technique can also be applied to evaluating frequency, stimulus probability, and frequency-weighted stimulus probability.

Syllable recognition probabilities p_s and phoneme recognition probabilities p_p, averaged across subjects, were calculated for each word and nonword target, and were used to calculate a *j* factor for each target using (3). Seventeen nonwords and two words were excluded because $p_s<0.05$. Eight words appeared twice in the word list and are represented only once in this and subsequent items analyses for a total of 110 word items, with p_p and p_s calculated over both sets of responses for the eight repeated target words. Figure 4 shows the resulting *j*-factor plot comparing words and nonwords, with average nonword $j=3.03\pm0.12$ and average word $j=2.41\pm0.14$, consistent with the subjects analysis. The *j*-factor model fits the items data less well than the subject data, with nonword RMSE=0.064 and word RMSE =0.088. The poorer fit is not unexpected since the entire word list is phonemically balanced while each word is obviously not phonemically balanced, resulting in a less uniform sample of p_p for the items analysis than for the subjects analysis.

Linear regressions were carried out on the word *j* factors, with frequency, stimulus probability, frequency-weighted stimulus probability, and the four density measures as the independent variables. The individual points are plotted in Fig. 5 with the best fitting lines for each independent variable. Table III presents r^2, *F*, *p*, and RMSE statistics for each regression. Only frequency ($p=0.048$), overall density

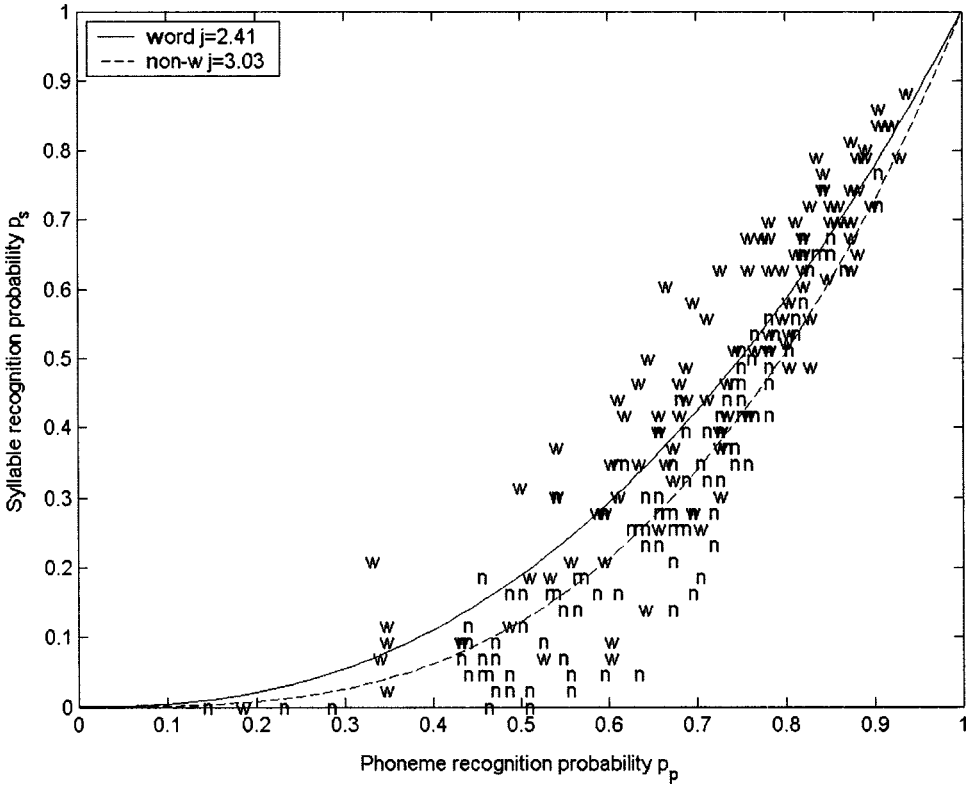


FIG. 4. Syllable recognition probability p_s as a function of phoneme recognition probability p_p for items. Each point represents the probabilities for a particular stimulus, word (w) or nonword (n), averaged over all subjects. The mean j factors, averaged over the points meeting the cutoff criteria for each set of stimuli, are plotted as $p_s = p_p^j$.

($p<0.001$), and CV density ($p<0.001$) are significantly correlated with the j factor. The CV density measure accounts for an extraordinarily large $r^2=0.623$ proportion of the variance in the j factors of different items and shows a low RMSE$=0.305$, both statistics indicating that CV density is well correlated with the j factor. Bootstrap analyses (10 000 samples) of the correlations between item j factors and the different density measures are consistent with the linear regression analyses, with only the distribution of CV density correlations not overlapping zero, while the distributions of VC density and CC density span both positive and negative values.

D. Discussion

In summary, nonword CVC stimuli were perceived as approximately $j=3$ independent units (with the units corresponding to phonemes) in analyses by both subjects and items, consistent with Boothroyd and Nittrouer's (1988) original findings. Words were found to be perceived as $j=2.34$ units in the analysis by subjects and $j=2.41$ in the analysis by items, also consistent with the previous study.

While the nonword j factors in this study and in Boothroyd and Nittrouer's (1988) study were both found to not differ significantly from 3, both studies found j factors slightly higher than 3. This result was also obtained in a subsequent study (Nittrouer and Boothroyd, 1990). The subsequent study also found slightly higher nonword j factors for geriatric listeners than for young adult listeners, and speculated that the geriatric listeners may have been more prone to impose lexical interpretations for nonword stimuli. The j factor slightly higher than 3 may be evidence that listeners are not neutral with respect to the nonwords but are in fact biased against that class of stimuli. While none of the three studies apparently has enough statistical power to verify that the j factors are significantly higher than 3, the consistency of the three studies is striking. The source of this effect is unknown.

A related issue is whether there are neighborhood effects contained within the nonword responses, and the validity of using the nonword responses in calculating neighborhood density measures. To the extent that the nonword j factor obtained was close to 3 (and in fact statistically does not differ from 3 in these data), support is provided for the assumption of independent phoneme recognition in nonwords, and the use of nonword responses to derive neighborhood density measures. The use of the nonword responses to derive neighborhood density measures is consistent with Luce and Pisoni (1998), who also used nonword response data to generate neighborhood density measures.

The new findings of the present study, then, consistent with the interpretation of the j factor as diagnostic of a bias in word recognition, are that frequency and neighborhood density are correlated with the j factor. High-frequency words and words in sparse neighborhoods show lower j factors than low-frequency words and words in dense neighborhoods. While stimulus probability does not appear to explain variance in the j factor in the items analysis, it was found to be a significant predictor of the j factor in the subjects analysis, with low stimulus probability words having a lower j factor than high stimulus probability words.

Considering the density results in more detail, the comparisons of effect sizes and the items analyses suggest that the advantage of CVC words in sparse neighborhoods is for the most part due to density defined by the initial two segments (CV) of overlap between competitors and the actual target.

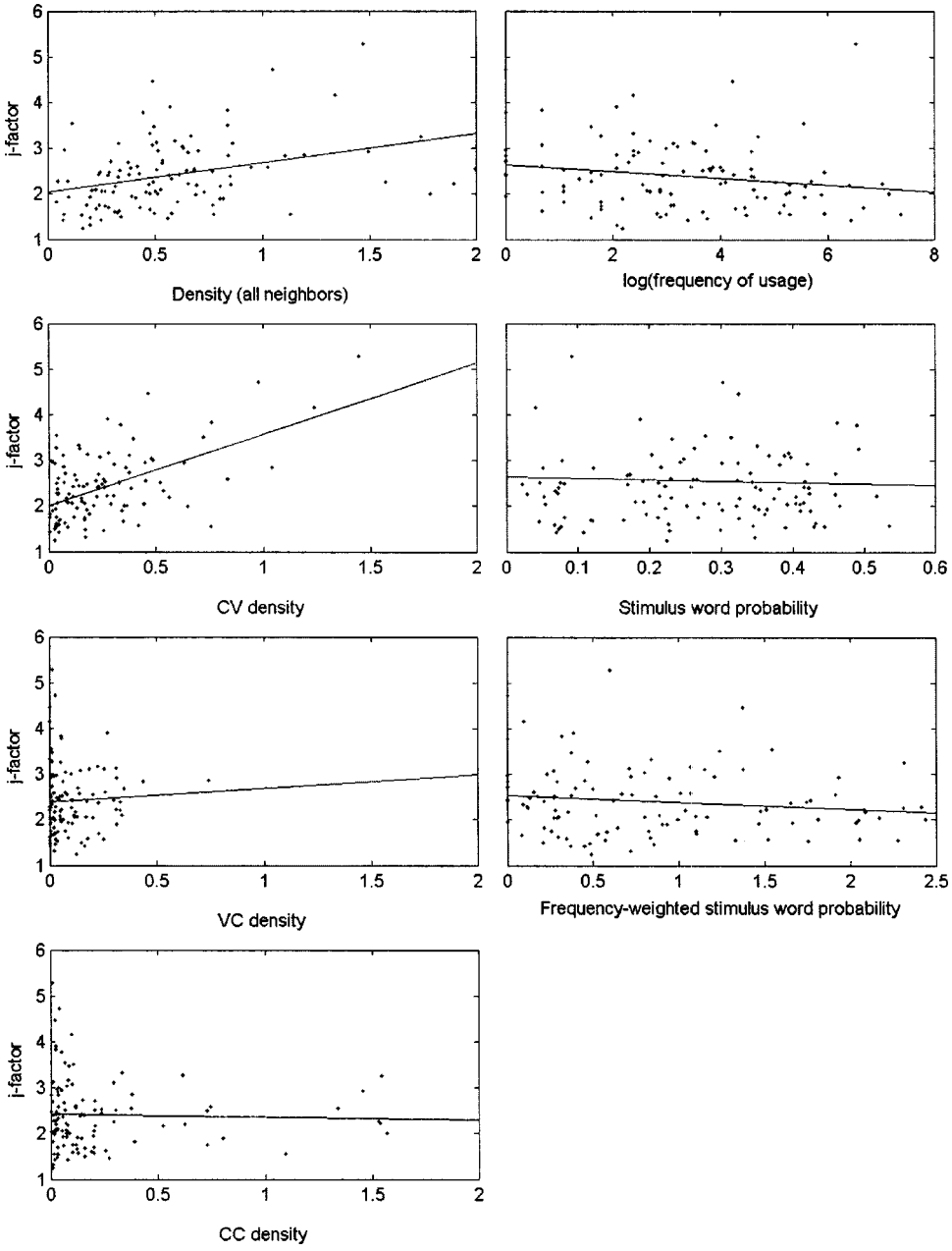


FIG. 5. Word *j* factors, averaged over subjects, as a function of overall density, CV density, VC density, CC density, log-frequency, stimulus probability, and frequency-weighted stimulus probability. The best-fitting lines from the linear regression analyses in Table III are plotted in each panel.

An implication of this latter result is that the final consonant (C_2) of words in sparse CV neighborhoods should show a higher recognition probability p_p than words in dense CV neighborhoods. Additionally, if the effect of VC and CC competitors is negligible, as suggested by the items analysis and the small effect found in the subjects analysis, neither VC nor CC density measures should be correlated with C_1 and V recognition probability, respectively (the segments that distinguish the target from its competitors for these density measures).

TABLE III. Goodness-of-fit statistics for linear regression models of the 110 item *j* factors.

Independent variable	r^2	F	p	RMSE
Density	0.123	15.17	<0.001	0.700
CV density	0.623	47.48	<0.001	0.305
VC density	0.002	0.27	0.604	0.747
CC density	<0.001	0.10	0.755	0.747
Log(frequency)	0.036	4.00	0.048	0.734
Stimulus probability	0.001	0.08	0.776	0.747
FW stimulus probability	0.019	2.11	0.149	0.740

It should be expected that the phoneme recognition probabilities p_p from the nonword confusion matrices should be highly correlated with the corresponding recognition probabilities p_p in the word responses. Accordingly, three multiple linear regressions on phoneme recognition probabilities were carried out to test these hypotheses. The average recognition probabilities p_p of C_1, V, and C_2 of each of the 110 words in the *j*-factor items analysis were modeled as functions of VC density and nonword p_p of C_1, CC density and nonword p_p of V, and CV density and nonword p_p of C_2, respectively. The results of these multiple regressions with partial correlation, as well as single regressions for both variables, are shown in Table IV. Figure 6 shows word C_1,

TABLE IV. Goodness-of-fit statistics for linear regressions of phoneme recognition probability p_p in 110 words. The dependent p_p values are the phoneme recognition probabilities in each word averaged over the 43 subjects at all S/N ratios. Two single-variable linear regressions are assessed for each dependent variable (nonword p_p and density) followed by the full model containing both variables. The partial correlations are for each independent variable in the full model. The independent nonword p_p probabilities are taken from the nonword confusion matrices.

Dependent	Independent	r^2	Partial corr.	F	df	p	RMSE
Word p_p of C_1	Nonword p_p of C_1	0.538	0.698	125.90	(1,108)	<0.001	0.133
	VC density	0.103	0.069	12.40	(1,108)	0.001	0.186
	Full model	0.540		62.90	(2,107)	<0.001	0.134
Word p_p of V	Nonword p_p of V	0.456	0.677	90.63	(1,108)	<0.001	0.118
	CC density	<0.001	−0.066	0.02	(1,108)	0.898	0.160
	Full model	0.459		45.32	(2,107)	<0.001	0.118
Word p_p of C_2	Nonword p_p of C_2	0.343	0.499	56.29	(1,108)	<0.001	0.176
	CV density	0.254	−0.383	36.68	(1,108)	<0.001	0.188
	Full model	0.439		41.9	(2,107)	<0.001	0.163

V, and C_2 phoneme recognition probabilities p_p as functions of the appropriate density measure and nonword p_p, as well as lines of best fit for single regressions for the individual independent variables.

These regression analyses confirm that the nonword phoneme recognition probability p_p is correlated with word p_p, as expected ($p<0.001$ in each case), but the results are also consistent with the hypothesis that only CV neighborhood density is negatively correlated with word C_2 phoneme recognition probability. The partial correlation for CV density in the multiple regression, $r=0.383$, is significant, indicating that when the intrinsic phoneme recognition probability is held constant, CV density inhibits C_2 p_p, or conversely, CV sparseness enhances $C_2 p_p$. A bootstrap analysis was consistent with the significant partial correlation for CV neighborhood density.

In the panel displaying word C_1 p_p as a function of VC density, a negative correlation is also apparent, and is confirmed with a bootstrap analysis. However, much of the effect may be due to an outlier (the word ''rot''), as confirmed by the negligible partial correlation $r=0.069$, and lack of improvement in rms error when is added to the model for nonword C_1 p_p. The CC density measure is not correlated at all with word V p_p, with a negligible partial correlation $r=-0.066$ and a flat slope in Fig. 6, consistent with a bootstrap analysis. These results are interpreted as consistent with listeners identifying C_2 of words in sparse CV neighborhoods at higher rates than words in dense CV neighborhoods, all other things (such as the phoneme in question) being equal.

The primacy of the CV density measure can be evaluated by comparing the performance of the neighborhood activation model (NAM) in (4) in predicting observed word syllable recognition probability p_s with versions of the neighborhood activation model that calculate neighborhood density using CV neighbors (CV-NAM), VC neighbors (VC-NAM), and CC neighbors (CC-NAM). If only CV density matters for the CVC stimuli used in this study, then there should be little difference between the full NAM and the CV-NAM, but the VC-NAM and CC-NAM should be worse at predicting word p_s values.

Following Luce and Pisoni's (1998) assessment of (4) with a correlation analysis in predicting word recognition probability, the responses at different S/N ratios were not averaged together. Instead, neighborhood densities and stimulus probabilities using the confusion matrices were computed for subjects at S/N ratios of −14, −11, −8, and −5 dB. Observed word syllable recognition probabilities p_s were similarly calculated by averaging across subjects at each S/N ratio for each word and S/N ratio combination, yielding 112 unique words×4 S/N ratios=448 points. Predicted p_s values were calculated for each word and S/N ratio for the four versions of the NAM, allowing for a different weight for the frequency-weighted neighborhood probability measure for each S/N ratio. Optimal weights were found with the Gauss–Newton nonlinear method with least squares, allowing the weight for each S/N ratio to vary to optimize the fit between the observed and predicted p_s values.

Table V shows the coefficients of correlation with nominal confidence intervals between the observed syllable recognition probabilities p_s and the different versions of the NAM with density weights, as well as log-frequency, frequency-weighted stimulus probability, and stimulus probability. Figure 7 shows observed p_s values plotted against the predicted values for the four versions of the NAM. The CV-NAM ($r=0.657$) and full NAM ($r=0.657$) are both highly correlated with observed p_s values, consistent with Luce and Pisoni (1998, Table I). For comparison, Luce and Pisoni (1998) report correlations of the NAM from $r=0.4043$ at +15 dB to $r=0.2277$ at −5 dB, based on a similar range of word recognition probabilities as the present study. They did not allow the effect of neighborhood density to vary with S/N ratio, but did analyze the data at different S/N ratios separately, obtaining higher correlations with the NAM at lower S/N ratios.

The correlation coefficients of the CV-NAM and NAM for the present data are virtually identical, consistent with the *j*-factor analyses indicating that the CV density measure is responsible for the vast majority of the density effect on CVC word recognition. The coefficients of correlation for the VC-NAM and CC-NAM are also significant but lower

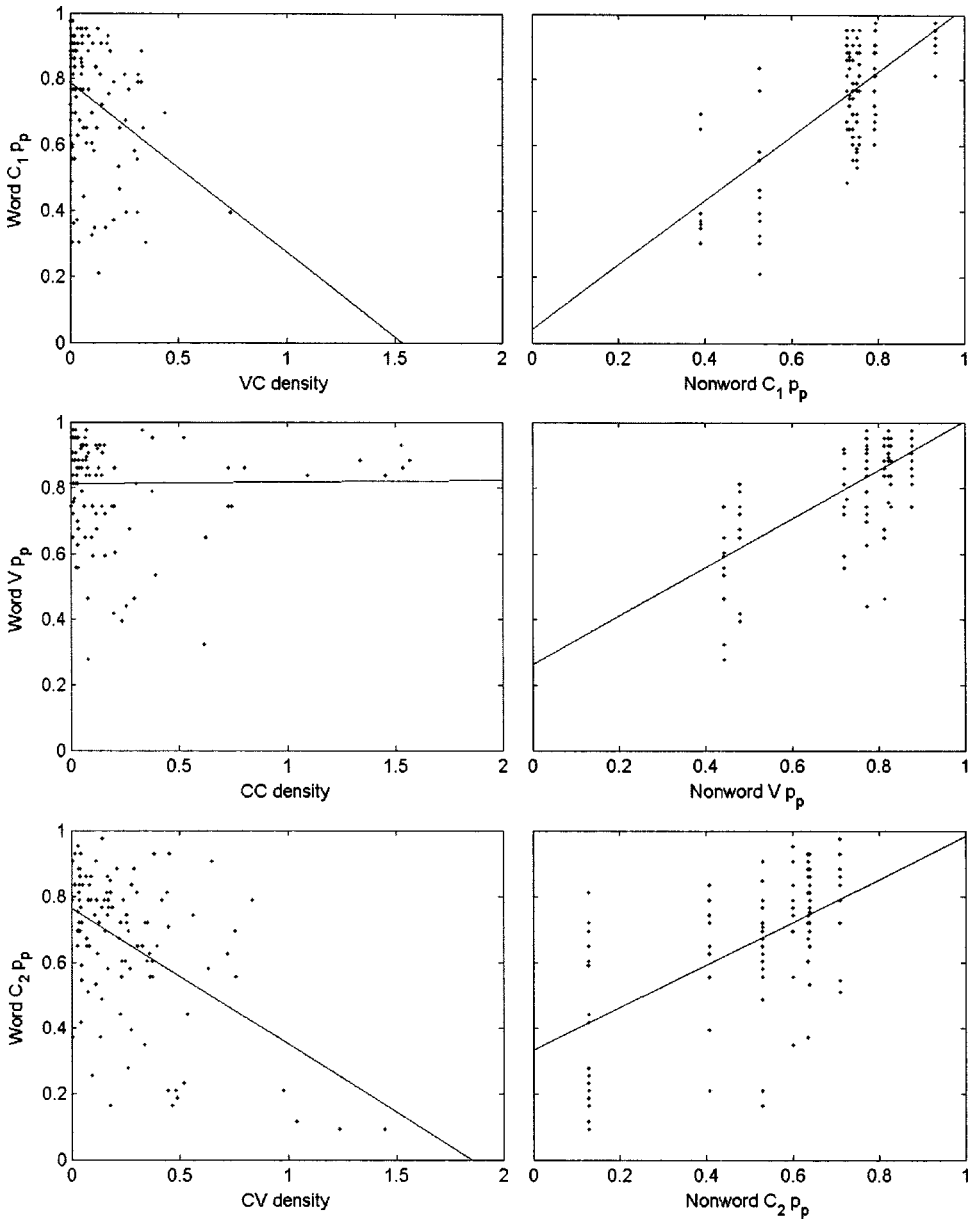


FIG. 6. Word phoneme recognition probabilities p_p, averaged over subjects, as a function of density and of nonword p_p. The upper panels show word p_p of C_1 as a function of VC density (left) and the nonword p_p taken from the C_1 confusion matrices, the middle panels show word p_p of V as a function of CC density and nonword p_p of V, and the bottom panels show word p_p of C_2 as a function of CV density and nonword p_p of C_2. The best-fitting lines from the single linear regression analyses in Table IV are plotted in each panel.

TABLE V. Correlations with nominal confidence intervals between observed word syllable recognition probability p_s and the predicted p_s values of the neighborhood activation models, frequency-weighted stimulus probability, log-frequency, stimulus probability, and CV-density adjusted stimulus probability. Each correlation is computed between 448 observed and predicted pairs (or values). Significance tests were computed by a transformation of the correlation into a t-statistic with df$=N-2=446$. All correlations are significant ($p<0.001$). The neighborhood density weights are also displayed for the neighborhood activation models for the -14, -11, -8, and -5 dB-S/N ratios.

Model	r	Nominal C. I.	-14 dB,	-11 dB,	-8 dB,	-5 dB weights
NAM	0.656	(0.599, 0.705)	3.470	1.716	0.714	0.421
CV-NAM	0.657	(0.601, 0.707)	14.154	6.265	1.675	1.096
VC-NAM	0.553	(0.485, 0.614)	44.277	23.920	4.475	3.928
CC-NAM	0.578	(0.513, 0.637)	43.652	26.211	1.549	0.314
FW stimulus Probability	0.591	(0.527, 0.648)	N/A			
Log-frequency	0.107	(0.014, 0.198)	N/A			
Stimulus probability	0.730	(0.684, 0.771)	N/A			
CV-density adjusted Stimulus probability	0.766	(0.725, 0.802)	N/A			

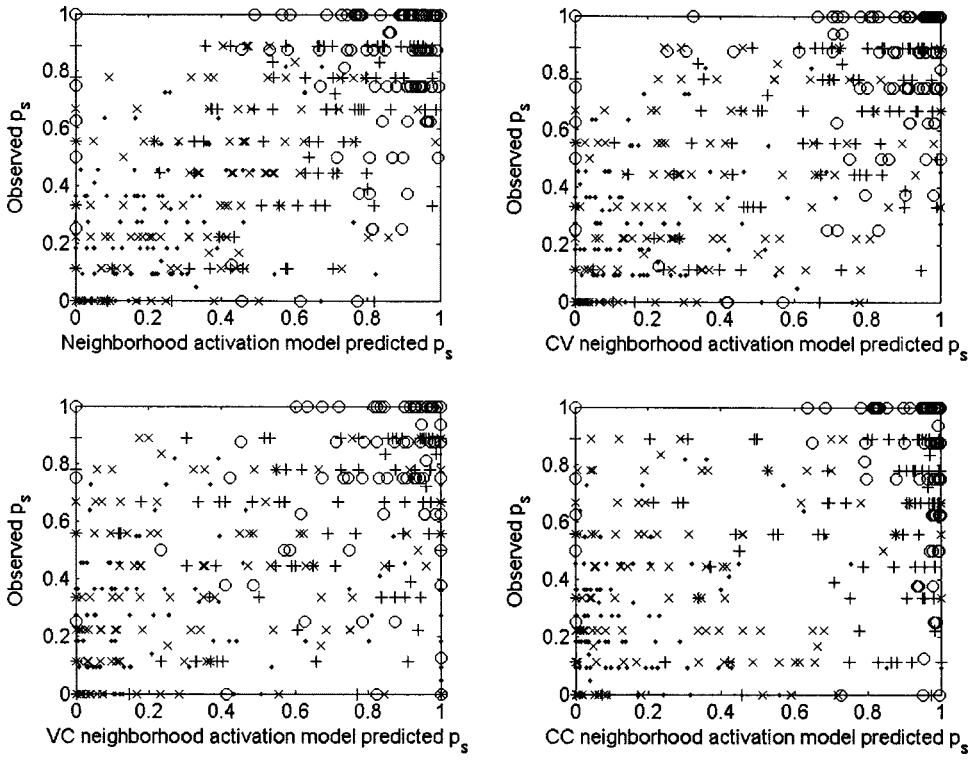


FIG. 7. Word syllable recognition probabilities p_s as predicted by the neighborhood activation models NAM, CV-NAM, VC-NAM, and CC-NAM. Data at −14 dB are points (·), at −11 dB crosses (×), at −8 dB plus signs (+), and at −5 dB circles (○).

than the CV-NAM and NAM values, with slightly overlapping confidence intervals. The neighborhood density weights for each version of the NAM are negatively correlated with S/N ratio, with the highest weights for the data at −14 dB and the lowest weights for the data at −5 dB, suggesting that the effects of neighborhood density are stronger in high noise conditions than in low noise conditions.

The significant correlation between observed p_s values and the predicted values of the CC-NAM and VC-NAM may be due to the presence of the frequency-weighted stimulus probability measure, which is identically present in all versions of the NAM evaluated above. The correlation analyses indicate a weak but significant correlation for log frequency, consistent with the literature. Frequency-weighted stimulus probability is highly correlated with observed p_s, explaining the performance of the VC-NAM and CC-NAM. More interesting is the extremely high correlation between the stimulus probability (not frequency-weighted) and observed p_s ($r = 0.730$). Luce and Pisoni (1998) do not report correlation between p_s and stimulus probability or frequency-weighted stimulus probability.

The stimulus probability measure, identical to (1) above, is essentially Fletcher's proposed starting point for understanding CVC word recognition in terms of nonsense CVC recognition. Boothroyd and Nittrouer's (1988) *k*-factor model of word CVC recognition as a function of nonsense CVC phoneme recognition may be useful

$$p_{\mathrm{WordP}} = 1 - (1 - p_{\mathrm{NonwordP}})^k, \tag{5}$$

where p_{WordP} and p_{NonwordP} are phoneme recognition probabilities in words and nonwords, respectively. Increases above 1 for the factor *k* quantify the extent of lexical status in improving phoneme recognition probability in words, and was found by Boothroyd and Nittrouer (1988) to be $k = 1.32 \pm 0.06$.

Boothroyd and Nittrouer apply (5) to model individual average word phoneme recognition probabilities as a function of their nonword recognition probabilities. The present results, particularly the analyses in Fig. 6 and Table IV, suggest that word C_1 and V recognition probabilities are well predicted by the corresponding phoneme recognition probabilities in nonwords. However, nonword C_2 recognition probability underpredicts word C_2 recognition probability, and that underprediction is inversely correlated with CV density. This inverse correlation can be implemented as an adjustment to *k* for each word S

$$k_{\mathrm{C2}} = \beta_0 + \beta_1 \sum_j p(CVN_j|S) \log(\mathrm{freq}_j), \tag{6}$$

where the summation term is CV density and the parameters β_1 and β_1 regulate the dependence on CV density. The parameter β_1 should be negative. The predicted phoneme recognition probability p_p for C_2 in a specific word can then be calculated using (5), and predicted syllable recognition probability p_s can be calculated using (1).

The Gauss–Newton method was used to calculate optimal values for β_0 and β_1 for each of the four CV density measures used in the above comparison of the different versions of the NAM. Predicted syllable recognition probabilities p_s for each of the 448 observed p_s values were calculated by taking the product of nonword p_p of C_1, V, and the k_{C2}-adjusted nonword p_p of C_2 using (5). This model is referred to as the *CV-density adjusted stimulus probability* model. Figure 8 shows the observed p_s values plotted against stimulus probability and the CV-density adjusted model, and Table VI shows the parameter values.

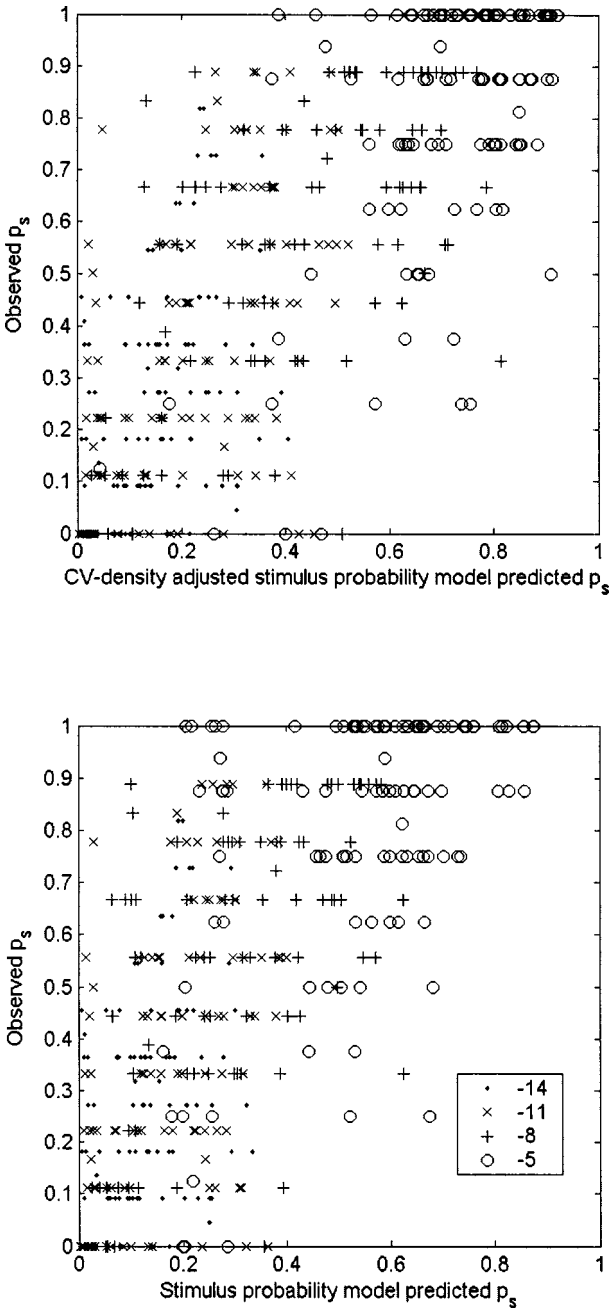


FIG. 8. Word syllable recognition probabilities p_s as predicted by the stimulus probability and CV-density adjusted stimulus probability models.

The CV-density adjusted stimulus probability model is the most highly correlated of all the models with observed syllable recognition probability ($r=0.766$). The model shows a slightly higher correlation than the stimulus probability model, indicating a slight improvement. The lower range of the confidence interval of the correlation coefficient overlaps with the upper boundary of the stimulus probability model, but not of that of the NAM or CV-NAM.

TABLE VI. Parameter values for CV-density adjusted stimulus probability model in (6).

Parameter	−14 dB	−11 dB	−8 dB	−5 dB
β_0	1.349	1.779	2.536	3.886
β_1	−0.057	−0.897	−1.370	−2.328

IV. GENERAL DISCUSSION

A. Interpretation of the *j* factor

The interpretation of the result $j=n$, as was found for nonsense CVC syllables with $n=3$ segments, is straightforwardly the confirmation of the hypothesis that the constituent segments of the syllables are perceived independently. But what does the result $j<n$ mean? Boothroyd and Nittrouer offer the interpretation that reduction of j below n measures the reduction in the number of independent perceptual units. Following their interpretation for the present study, words are perceived as consisting of $j=2.34$ independent units, with each phoneme consisting of about 0.78 units.

As discussed above, Nearey (2001, in press), proceeding from a computational simulation of Boothroyd and Nittrouer's results, suggests that small reductions (around 1 or less for $n=3$) in the j factor could arise from a bias in favor of particular items in a Luce choice model of word recognition. Under this interpretation, results of $2<j\leqslant n$ are consistent with independent perception of n segments, and reduction of j below n quantifies the amount of bias involved for those items. The present result for word frequency, that high-frequency words have lower j factors than low-frequency words, and neighborhood density, that words in dense neighborhoods have higher j factors than words in sparse neighborhoods are in line with this interpretation, and the widely held view that word frequency and neighborhood density influence word recognition as biases.

An alternative possibility, but not necessarily mutually exclusive with the bias interpretation, is that the high-frequency words were produced in a more casual and coarticulated style than the low-frequency words, resulting in a phonemic interdependence with a reduced j factor. Although clear speech is more intelligible (Picheny *et al.*, 1985), an acoustic-phonetic investigation by Bradlow (in press) indicates that clear and casual speech contain equal amounts of CV coarticulation. In the present study, degree of coarticulation in the stimuli was not controlled for, so it may be difficult to assess the viability of this explanation, but a future investigation might pursue this issue.

An important issue is the generalizability of the present results on English CVC syllable identification to the recognition of longer words in English and to general word recognition in other languages. With respect to longer words in English, it is likely that the j factor would be much less than n segments for words of length n, reflecting the greater uniqueness of longer words. The relationship of the present results to other languages will depend strongly on the possible syllable structures. For example, CVC syllables are the most common type of syllables in English, while other languages place strong restrictions on syllable-final consonants (e.g., Spanish and Japanese). Investigations of these other languages using the present methodology would provide an interesting opportunity to examine the issue of density and temporal order, since words in these languages may have later uniqueness points as a result of the lack of syllable-final consonants.

B. Temporal restrictions on neighborhood density effects

The neighborhood density results suggest that the mechanism by which the bias is used by listeners is a function of the temporal order of the acoustic-phonetic information in the speech signal. Recall that, as predicted, words with sparse neighborhoods had a lower *j* factor than words with dense neighborhoods, consistent with a bias favoring words from sparse neighborhoods. Importantly, this result largely holds true of neighborhoods in which the target and neighbors share the first two segments and somewhat for those sharing the first and last segments (though opposite expectation, and only in the subjects analysis), but not the neighborhoods in which the target and neighbors share the last two segments. These findings are consistent with a recent investigation of reaction times in auditory shadowing and lexical decision (Vitevitch, 2002). In that study, slight but significant reaction time advantages (10 ms difference between RT means of approximately 1000 ms) were found for one-syllable target words whose phonological neighbors predominantly shared final phonemes rather than initial phonemes. Additonally, results from reaction time and phoneme-position investigations of the lexical identification shift (Ganong, 1980) are also consistent with the present temporal effect findings in word recognition. In those lexical identification shift studies, elements of a nonword-word stimulus continuum are more likely to be identified as words for late response times rather than early response times (Fox, 1984). Although there are results showing the effect with the ambiguous phoneme in word-initial position, it is also reported to be more reliable when the ambiguous phoneme occurs later in the word (Pitt and Samuel, 1993).

The lack of an effect of VC density on the *j* factor indicates that in open-response identification, contextual information from correct recognition of syllable-final material cannot bias the perception of earlier-occurring syllable-initial material in the same way that bias appears to affect the perception of upcoming material (cf. Salasoo and Pisoni, 1985). If the role of bias were merely to narrow such a set, along the lines of what Broadbent (1967) calls the sophisticated guessing model, then one would expect a significant reduction of the *j* factor for words in sparse VC neighborhoods as well.

Correct perception of the beginning of a word in a sparse phonetic neighborhood delimits a sufficiently small set of potential candidates. A listener can then focus attention on just those phonetic features that distinguish the members of this small set to achieve correct recognition despite reduced acoustic-phonetic information present at the end of the syllable. This account of how listeners use bias, based on a temporal analysis of the effects of neighborhood density, offers support for the dynamic aspects (but perhaps not the strict autonomy) of the cohort theory of word recognition (Marslen-Wilson, 1989). In the cohort theory, the beginning of a stimulus defines a candidate set, which is pruned until the isolation point, when the acoustic-phonetic evidence uniquely identifies the target from within that set.

Alternatively, the lack of any significant *j*-factor effect for VC neighborhood density could be because C_2 is less intelligible than C_1 (see Wright, 2001 for original data and extensive review of the increased intelligibility of word-initial consonants relative to word-final consonants), resulting in a reduced range for VC density relative to CV density, as reflected in the histogram of neighborhood density measures in Fig. 3. While a negative result cannot disprove a hypothesis, the lack of *any* effect for the VC or CC density measures in the linear regressions of the items analysis is noteworthy. It is not likely that reduced intelligibility of C_2, with concomitant reduced range of VC density, is the *only* explanation for a lack of a VC density effect. The virtually identical correlations between word syllable recognition probability p_s and both the NAM and CV-NAM, along with significantly less correlation with the VC-NAM and CC-NAM, are consistent with a temporal order constraint on density effects.

C. Interaction between the lexicon and the sensory channel

Frequency of usage and neighborhood density effects could be explained here in terms of a criterion bias shift, consistent with noninteractive (feedforward) models of top-down effects in word recognition such as Merge (Norris, McQueen, and Cutler, 2000) or the fuzzy logical model of perception (FLMP; Massaro and Cohen, 1991). The high correlation between observed syllable recognition probability and the present implementation of the (CV-) NAM in (4), which explains word frequency and neighborhood density effects as biases in a Luce choice model, is consistent with the noninteractive approach.

However, the fact that the CV-density adjusted stimulus probability model is the most highly correlated with observed syllable recognition probability poses a problem for noninteractive theories of word recognition. In the CV-density adjusted stimulus probability model, the phoneme recognition probability p_p of C_2 of words in sparse neighborhoods is increased *not* by a bias, but as an explicit interaction between the lexicon and the sensory channel. This approach may be more consistent with interactive (feedback) models of word recognition such as TRACE (McClelland and Elman, 1986) or adaptive resonance theory (Grossberg and Stone, 1986; see also Grossberg, 2000; Luce *et al.* 2000), in which contextual information directly influences the sensory channel at a relatively lower level of processing.

The difference in correlation values, though significant, between the NAM and the CV-density adjusted stimulus probability model may diminish with more sophisticated modeling techniques, such as more realistic representations of sensory input. However, it may turn out that resolution of the interactive/noninteractive issue in word recognition is not possible given the focus of most of the current approaches on relatively higher-level processes. None of the aforementioned models of word recognition (FLMP, Cohort, TRACE, adaptive resonance theory, or Merge) begins with realistic raw acoustic input, but instead starts from some phonological or phonetic feature time-ordered representation to produce predictions of reaction time and accuracy data in different

contextual conditions (but cf. Johnson, 1997). In this respect, the present approach of modeling word syllable recognition probability on the basis of nonsense phoneme recognition probability is not much better, given that phoneme recognition probabilities are only slightly less abstract than the phonemes themselves.

Certainly these research programs have contributed much to our understanding of word recognition, and they still have much to provide. However, if the mechanisms of relatively lower-level processes such as phoneme or phonological feature identification impose critical but as yet unknown constraints on more central lexical access processes—even if the interface between the levels is well defined—then the progress of current approaches in understanding the central processes may become limited at some point. If this is the case, then it becomes imperative for research on lexical access to take seriously the problems of speech invariance beyond limited consideration of contextual variability. It is worthwhile to note that Stevens's (2002) recent proposal of a research program to provide a model of lexical access is primarily concerned with detailing what is known about how phonological features are extracted from the speech signal, and that the solution of this problem is necessary for a complete understanding of the whole process of lexical access.

V. CONCLUSION

Support has been provided for Boothroyd and Nittrouer's *j* factor as a robust and replicable measure of the effects of context in human speech perception. The *j* factor represents the number of perceptually independent parts within a whole, and can be interpreted as a bias in favor of words over nonsense syllables, words with higher usage frequencies over words with lower usage frequencies, or of words from sparse phonetic neighborhoods over words from dense neighborhoods. The neighborhood density effect has temporal restrictions, such that the neighborhood is primarily determined by the beginning of a CVC word. This temporal effect appears to improve perception of the final segments of CVC words when listeners can take advantage of a sparse neighborhood, as demonstrated in a *k*-factor model incorporating CV neighborhood density effects to predict phoneme recognition probability of the word-final consonant.

ACKNOWLEDGMENTS

Special thanks to Katherine Cassady and Sherene Flemmings for running the experiment. I am grateful to audience members, particularly Pam Beddor, Julie Boland, and John Kingston for insightful comments. Reviews from Arthur Boothroyd, Terry Nearey, and an anonymous reviewer improved this paper enormously. All errors are my own.

APPENDIX:

These are the stimuli used by Boothroyd and Nittrouer and in the present study. Note that in the original Boothroyd and Nittrouer word list, eight words were repeated, namely *bag, bed, beg, cause, like, rule, ten,* and *time*. The distribution of the original stimuli was maintained in the present study by recording two versions of these items. About 24 of the 120 original Boothroyd and Nittrouer nonsense syllable stimuli may be actual lexical items for some speakers, either because of an /ɑ/–/ɔ/ merger (e.g., /sɑk/ *sock*), the items may be familiar proper names (e.g., /lɛn/ *Len*), or in the case of one item, the possibility of a morphologically complex word was not considered (e.g., /kiz/ *keys*). These items were kept in the present study.

CVC words:

bag	dial	lean	pal	says
ball	dies	let	pass	seal
bed	dike	lice	peace	sews
beg	dip	like	peck	sick
bell	don	load	pen	sip
big	dot	loan	pep	soap
boom	doze	log	pick	soon
boss	hag	loon	pies	sop
buys	hall	lose	pig	sought
call	ham	mace	pope	suit
cat	heat	made	race	tag
caught	heel	maim	rake	take
cause	hem	make	rice	tame
coat	hid	mall	rid	tell
cod	hide	mean	rig	ten
comb	hiss	meek	rod	time
come	hog	mid	room	toes
cop	hoop	miss	root	tomb
dad	hope	mock	rose	toss
date	keep	moon	rot	tote
dawn	keys	mop	rule	
dean	laws	pack	same	
deep	league	pad	sass	

CVC nonsense syllables:

maɪg	beɪp	kɛz	bɪg	ruk
kɔs	raik	pɪd	tæt	teɪt
heɪm	lim	sum	pɛm	laɪl
dɪt	mɪn	raɪg	roʊs	bɔp
tup	put	teɪn	kɑl	pɪm
bɑk	soʊg	loʊl	hun	kæz
poʊd	tæz	mɑt	dɪz	doʊn
sɛm	dɑs	dɔk	leɪp	hɑd
læl	hɔd	bis	sɔk	mis
riz	kɛl	hæp	maɪd	sɛg
rit	reɪg	lɛn	pæz	reɪm
seɪz	tuk	dæp	his	lɔk
bɔd	mɔt	paɪt	mɛk	mul
kum	kɪp	bɪm	soʊm	dæs
lɛp	boʊd	sig	tɑl	hɛz
poʊn	hɑs	hɑl	lɔt	tid
haɪs	læn	keɪz	keɪd	poʊt
dæk	saɪl	rus	baɪp	bɑn
tig	dɛz	tɔd	dug	kaɪp
mɑl	pim	moʊk	rɪn	sɪg
puk	kiz	teɪd	dɔs	haɪm
lit	sæn	bɛp	moʊp	rɑl
kɪg	peɪk	taɪs	dɔd	miz
rɛm	hæp	lɑn	soʊt	bul

Allen, J. (**1994**). ''How do humans process and recognize speech?'' IEEE Trans. Speech Audio Process. **2**, 567–577.
Beddor, P. S., and Evans-Romaine, D. (**1995**). ''Acoustic-perceptual factors in phonological assimilations: A study of syllable-final nasals,'' Rivista di Linguistica **7**, 145-174.
Benkí, J. R. (**to appear**). ''Analysis of English nonsense syllable recognition in noise,'' Phonetica.
Boothroyd, A., and Nittrouer, S. (**1988**). ''Mathematical treatment of context effects in phoneme and word recognition,'' J. Acoust. Soc. Am. **84**, 101–114.
Bradlow, A. (**in press**). ''Confluent talker- and listener-oriented forces in clear speech production,'' in *Papers in Laboratory Phonology VII*, edited by C. Gussenhoven, T. Rietveld, and N. Warner (Cambridge Univerisity Press, Cambridge).
Broadbent, D. E. (**1967**). ''Word-frequency effect and response bias,'' Psychol. Rev. **74**, 1–15.
Fox, R. (**1984**). ''Effect of lexical status on phonetic categorization,'' J. Exp. Psychol. Hum. Percept. Perform. **10**, 526–540.
Fletcher, H. (**1953**). *Speech and Hearing in Communication* (Van Nostrand, New York).
Ganong, W. F. (**1980**). ''Phonetic categorization in auditory word perception,'' J. Exp. Psychol. Hum. Percept. Perform. **6**, 110–125.
Goldinger, S. D. (**1997**). ''Words and voices: Perception and production in an episodic lexicon,'' in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic, San Diego), pp. 33–66.
Grossberg, S., and Stone, G. (**1986**). ''Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance,'' Psychol. Rev. **93**, 46–74.
Grossberg, S. (**2000**). ''Brain feedback and adaptive resonance in speech perception,'' Behav. Brain Sci. **23**, 332–333.
Hura, S. L., Lindblom, B., and Diehl R. L. (**1992**). ''On the role of perception in shaping phonological assimilation rules,'' Lang. Speech **35**, 59–72.
Johnson, K. (**1997**). ''The auditory/perceptual basis for speech segmentation,'' in *OSU Working Papers in Linguistics No. 50, Papers from the Linguistics Laboratory*, edited by K. Ainsworth-Darnell and M. D.'Imperio (Ohio State University, Department of Linguistics, Columbus), pp. 101–113.
Kucera, F., and Francis, W. (**1967**). *Computational Analysis of Present Day American English* (Brown University Press, Providence).
Lenzo, K. (**1998**). ''T2P: Text-to-phoneme converter builder software,'' retrieved from http://www-2.cs.cmu.edu/~lenzo/t2p/
Luce, P. A., and Pisoni, D. B. (**1998**). ''Recognizing spoken words: The neighborhood activation model,'' Ear Hear. **19**, 1–36.
Luce, P. A., Goldinger, S. D., and Vitevitch, M. S. (**2000**). ''It's good...but is it ART?'' Behav. Brain Sci. **23**, 336.
Marslen-Wilson, W. (**1989**). ''Access and integration: Projecting sound onto meaning,'' in *Lexical Representation and Process*, edited by W. Marslen-Wilson (MIT Press, Cambridge), pp. 3–24.
Massaro, D. W. (**1998**). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle* (MIT Press, Cambridge).
Massaro, D., and Cohen, M. (**1991**). ''Integration versus interactive activation: The joint influence of stimulus and context in perception,'' Cognit Psychol. **23**, 558–614.
McClelland, J. L., and Elman, J. L. (**1986**). ''The TRACE model of speech perception,'' Cognit. Psychol. **18**, 1–86.
Miller, G. A., and Nicely, P. E. (**1955**) ''An analysis of perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–352.
Nearey, T. N. (**1990**). ''The segment as a unit of speech perception,'' J. Phonetics **18**, 347–373.
Nearey, T. N. (**2001**). ''Phoneme-like units and speech perception,'' Lang. Cognit. Proc. **16**, 673–681.
Nearey, T. N. (**in press**). ''On the factorability of phonological units in speech perception,'' in *Phonetic Interpretation: Papers in Laboratory Phonology IV*, edited by J. Local, R. Ogden, R. Temple (Cambridge University Press, Cambridge).
Nittrouer, S., and Boothroyd, A. (**1990**). ''Context effects in phoneme and word recognition by young children and older adults,'' J. Acoust. Soc. Am. **87**, 2705–2715.
Nooteboom, S. (**1968**) ''Perceptual confusions among Dutch vowels presented in noise,'' IPO Annual Progress Report, vol. 3 (Instituut voor Perceptie Onderzoek, Eindhoven), pp. 68–71.
Norris, D. G. (**1986**). ''Word recognition: Context effects without priming,'' Cognition **22**, 93–136.
Norris, D. G., McQueen, J. M., and Cutler, A. (**2000**). ''Merging information in speech recognition: Feedback is never necessary,'' Behav. Brain Sci. **23**, 299–370.
Nusbaum, H. C., Pisoni, D. B., and Davis, C. K. (**1984**). ''Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20 000 words,'' *Research on Speech Perception Progress Report No. 10* (Speech Research Laboratory, Psychology Department, Indiana University, Bloomington).
Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). ''Speaking clearly for the hard of hearing. I. Intelligibility differences between clear and conversational speech,'' J. Speech Hear. Res. **29**, 434–446.
Pickett, J. M. (**1957**). ''Perception of vowels heard in noises of various spectra,'' J. Acoust. Soc. Am. **29**, 613–620.
Pitt, M. A., and Samuel, A. G. (**1993**). ''An empirical and meta-analytic evaluation of the phoneme identification task,'' J. Exp. Psychol. Hum. Percept. Perform. **19**, 699–725.
Salasoo, A., and Pisoni, D. B. (**1985**). ''Interaction of knowledge sources in spoken word recognition,'' J. Memory Lang. **24**, 210–231.
Savin, H. (**1963**). ''Word-frequency effects and errors in the perception of speech,'' J. Acoust. Soc. Am. **35**, 200–206.
Schroeder, M. R. (**1968**). ''Reference signal for signal quality studies,'' J. Acoust. Soc. Am. **44**, 1735–1736.
Stevens, K. N. (**2002**). ''Toward a model for lexical access based on acoustic landmarks and distinctive features,'' J. Acoust. Soc. Am. **111**, 1872–1891.
Vitevitch, M. (**2002**). ''Influence of onset density of spoken-word recognition,'' J. Exp. Psychol. Hum. Percept. Perform. **28**, 270–278.
Wang, M. D., and Bilger, R. C. (**1973**). ''Consonant confusions in noise: A study of perceptual features,'' J. Acoust. Soc. Am. **54**, 1248–1266.
Webster's Seventh Collegiate Dictionary (**1967**). (Library Reproduction Service, Los Angeles).
Wright, R. (**2001**). ''Perceptual cues in contrast maintenance,'' in *The Role of Speech Perception in Phonology*, edited by E. Hume and K. Johnson (Academic, San Diego), pp. 251–277.

The effects of hearing loss on the contribution of high- and low-frequency speech information to speech understanding[a)]

Benjamin W. Y. Hornsby[b)] and Todd A. Ricketts
Dan Maddox Hearing Aid Research Laboratory, Vanderbilt Bill Wilkerson Center, 1114 19th Avenue South, Nashville, Tennessee 37212

(Received 24 May 2002; revised 13 December 2002; accepted 17 December 2002)

The speech understanding of persons with ''flat'' hearing loss (HI) was compared to a normal-hearing (NH) control group to examine how hearing loss affects the contribution of speech information in various frequency regions. Speech understanding in noise was assessed at multiple low- and high-pass filter cutoff frequencies. Noise levels were chosen to ensure that the noise, rather than quiet thresholds, determined audibility. The performance of HI subjects was compared to a NH group listening at the same signal-to-noise ratio and a comparable presentation level. Although absolute speech scores for the HI group were reduced, performance improvements as the speech and noise bandwidth increased were comparable between groups. These data suggest that the presence of hearing loss results in a uniform, rather than frequency-specific, deficit in the contribution of speech information. Measures of auditory thresholds in noise and speech intelligibility index (SII) calculations were also performed. These data suggest that differences in performance between the HI and NH groups are due primarily to audibility differences between groups. Measures of auditory thresholds in noise showed the ''effective masking spectrum'' of the noise was greater for the HI than the NH subjects. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1553458]

PACS numbers: 43.71.Ky, 43.66.Ts, 43.66.Sn [CWT]

I. INTRODUCTION

Previous research suggests that the presence of sensorineural hearing loss (SNHL) may reduce the contribution of speech information in a given frequency region to speech understanding (i.e., Pavlovic *et al.*, 1986; Studebaker *et al.*, 1997). What is not clear, however, is whether SNHL has a differential effect on the contribution of speech information depending on the frequency region where the hearing loss occurs. At least three differing viewpoints on this question have been expressed in the literature.

First, early work in this area suggests that the presence of hearing loss results in a ''uniform deficit'' in the contribution of speech information across all affected frequencies (Boothroyd, 1978; Pavlovic, 1984; Pavlovic *et al.*, 1986; Studebaker *et al.*, 1997). In a 1978 paper, Boothroyd discussed a series of experiments in which he measured speech understanding under various conditions of low- and high-pass filtering. He used these results to determine the relative contribution of different frequency regions to phoneme identification for children with hearing loss (Boothroyd, 1967, 1968). Study results showed that for persons with flat hearing losses the contribution of speech information was reduced across all frequencies equally. In contrast, individuals with high-frequency losses showed a reduced contribution of speech information primarily in the regions where hearing loss was present and followed the normal pattern in regions where hearing was near normal. In other words, he observed that it was primarily the **presence** of hearing loss that resulted in a reduction in the contribution of a frequency region to speech understanding, regardless of the frequency region where the loss occurred.

Other support for a ''uniform deficit'' comes from earlier work utilizing the articulation index (ANSI S3.5, 1969) or AI (now referred to as the speech intelligibility index or SII, ANSI S3.5, 1997) to investigate deficits in speech understanding of persons with hearing loss not explained by reduced audibility or adverse listening conditions such as high presentation levels. Several researchers developed correction factors to account for the negative effects of hearing loss on speech understanding (Pavlovic, 1984; Pavlovic *et al.*, 1986; Studebaker *et al.*, 1997) and have reported good improvements in their predictive accuracy. The magnitude of these correction factors, however, was independent of the frequency where the hearing loss occurred.

Data from several recent studies provide a contrasting view on the impact of hearing loss on speech information in various frequency regions. These studies suggest that hearing loss may result in a ''frequency-specific'' deficit in the contribution of speech information. Specifically, persons with hearing loss may be less able to make use of amplified high-frequency information (i.e., above 3000 Hz) than amplified low-frequency information, particularly when their thresholds are worse than 55–80 dB HL (Ching *et al.*, 1998; Hogan and Turner, 1998; Turner and Cummings, 1999; Amos, 2001). Hogan and Turner (1998) described the ''efficiency'' with which persons with high-frequency sloping SNHL were able to make use of speech information in various frequency regions. They found that the persons with hearing loss were limited in their ability to make use of amplified speech information above 4000 Hz, particularly when

[a)]Portions of this research were presented in a platform paper presented at the American-Speech-Language-Hearing Association National Convention, New Orleans, LA, November 2001, and during a poster session at the American Auditory Society Spring meeting, Scottsdale, AZ, March 2002.
[b)]Electronic mail: ben.hornsby@vanderbilt.edu

 0001-4966/2003/113(3)/1706/12/$19.00

the degree of hearing loss in this frequency region exceeded about 55 dB HL. This group appeared better able to make use of lower frequency information even in the presence of a similar degree of hearing loss.

Ching *et al.* (1998, 2001), using slightly different methods and participants with a wider range of hearing losses, reported similar findings. These authors compared the sentence recognition performance of persons with hearing loss, using filtered sentence materials in quiet, to SII predictions. The authors derived several correction factors to the standard SII procedure in an attempt to reduce the error between predicted and observed scores. The best fit occurred when ''individual frequency-dependent proficiency'' corrections were applied in conjunction with the standard SII correction for high presentation levels. The frequency-dependent correction factors were largest in the high-frequency regions.

Turner and Cummings (1999) reported findings similar to Hogan and Turner and Ching *et al.* Their study participants, primarily persons with high-frequency SNHL, listened to unfiltered nonsense syllables in quiet as presentation levels were systematically increased until asymptotic performance levels were reached. Their results also suggested that restoring the audibility of high-frequency information (above about 3000 Hz) to persons with high-frequency hearing losses greater than about 55 dB HL provided limited benefits in terms of speech understanding.

A third suggestion in the literature is that persons with hearing loss differ in their ability to make use of amplified speech information in various frequency regions and that these differences may be due to the presence or absence of ''cochlear dead regions'' (Moore *et al.*, 2000; Vickers *et al.*, 2001; Baer *et al.*, 2002). A dead region has been defined as a region of the basilar membrane associated with a ''complete loss of inner hair cells'' (Moore *et al.*, 2000). Vickers *et al.* (2001) compared the speech understanding in quiet of two groups of persons with hearing loss at multiple low-pass filter cutoff frequencies. One group of subjects had high-frequency dead regions while the other group did not. Potential dead regions within the cochlea were identified using both psychophysical tuning curves and the threshold equalizing noise (TEN) test (Moore *et al.*, 2000). The TEN test is a clinical test that measures auditory thresholds in quiet and in a spectrally shaped broadband noise. Thresholds in noise that are abnormally elevated (as described later in the text) are suggestive of dead regions. Vickers *et al.* (2001) found that, in general, individuals with identified high-frequency dead regions made limited use of amplified high-frequency information to improve speech understanding. In contrast, subjects with hearing loss but without dead regions showed more consistent improvement in speech understanding as low-pass filter cutoff frequency was progressively increased. Baer *et al.* (2002) reported a similar finding for subjects with and without dead regions listening to nonsense syllables in a noise background.

The research discussed here suggests that our understanding of the effects of hearing loss on the contribution of high-frequency speech information remains unclear. Recent findings suggesting a frequency-specific deficit focused primarily in the high frequencies have significant implications in terms of defining appropriate amplification requirements for persons with hearing loss and as such warrant serious consideration. For example, based in part on this research, Ching *et al.* (2001) suggested that to achieve maximum ''effective audibility'' only minimal or even no gain should be provided in some high-frequency regions in the presence of severe or greater hearing loss, so that more gain may be provided to regions with less hearing loss. Following this suggestion would ensure that certain high-frequency speech sounds would remain essentially inaudible for those listeners with severe high-frequency hearing loss. Given the significant implications of the recent findings in this area, further research appears warranted.

This study further investigates the role hearing loss plays in limiting the contribution of speech information in various frequency regions. Specifically, to limit the confound of degree of hearing loss across frequency, subjects with flat losses are used to examine the impact of SNHL on the contribution of high-and low-frequency information to speech understanding. In addition, if frequency-specific deficits are observed, a second purpose of this study is to determine if these deficits are related to the presence of cochlear dead regions.

II. METHODS

A. Participants

A total of 27 participants, 18 with normal hearing and 9 with hearing loss, participated in this study. All participants with normal hearing passed a pure-tone air conduction screening at 20 dB HL (250–8000 Hz; ANSI S3.6, 1996) and had no history of otologic pathology. Participants with normal hearing were randomly divided into two groups (nine per group). One group, identified as the normal-hearing unshaped group (NHU), was used to obtain baseline performance on the speech recognition task used in this study. Participants in this group (two male, seven female) ranged in age from 19 to 31 years (mean 25.4). The other group, normal-hearing shaped (NHS), listened to speech that was spectrally shaped and provided a control group for the individuals with hearing loss which also listened to shaped speech. Speech shaping for listeners with normal and impaired hearing is described later. Individuals in the NHS group (three male, six female) ranged in age from 20 to 37 years (mean 27.2).

Nine persons with flat, moderate-to-severe SNHL (HI group) also participated in this experiment. Flat hearing loss was operationally defined as auditory thresholds ranging between 55 and 70 dB HL at octave frequencies of 500–4000 Hz. The slope of hearing loss, defined as the difference in thresholds at 500 and 4000 Hz, across participants was −5.6 dB, with seven of the nine participants having a slope of 0 to −5 dB.

Participants with hearing loss ranged in age from 42 to 80 years old (mean 66.4 years). Specific demographic details of the persons with hearing loss are provided in Table I. Several subjects reported that their hearing loss was due primarily to noise exposure and aging. Although a flat configuration is not typically associated with these factors, no other

TABLE I. Demographic characteristics of the participants with hearing loss. HL: hearing loss; HA: hearing aids.

Subject	Sex	Age	Length of HL (in years)	HA use (in years)	Binaural aids	Education	Cause of HL
DJ	M	76	20	18	Y	14	Noise/Presbycusis
JA	M	80	20	20	Y	16	Noise/Presbycusis
MB	M	80	22	15	Y	14	Noise/Presbycusis
DG	F	70	12	5	N	12	Presbycusis
JM	M	80	20	20	Y	14	Noise/Presbycusis
SG	F	42	40	30	Y	18	Congenital
AL	F	72	5	2	Y	15	Presbycusis/Genetic
VP	F	47	16	6	Y	12	Unknown
SP	M	51	15	15	Y	16	Noise/Unknown
Average		66.4	18.9	14.6		14.6	

relevant factors were reported in the subject's history.

The auditory thresholds of participants with hearing loss were assessed at octave frequencies between 250 and 8000 Hz, as well as at the interoctave frequencies of 3000 and 6000 Hz (ANSI S3.6, 1996). Participants exhibited essentially symmetrical hearing loss (interaural difference of $\leqslant$20 dB); air-bone gaps $\leqslant$10 dB at all frequencies, and other than hearing loss, reported no history or complaints of otologic pathology, surgery, or unilateral tinnitus. All testing, following the initial threshold assessment, was performed monaurally. The ear with the smallest difference in thresholds between 500 and 4000 Hz was chosen for testing. In contrast, the test ear was systematically varied for the NHU and NHS groups. Table II lists the auditory thresholds, of the ear tested, for each of the HI participants.

B. Procedures

Sentence recognition in noise, at various filter cutoff frequencies, was assessed for all three groups (NHU, NHS, and HI). In addition, the NHS and HI groups also completed threshold testing in a speech-shaped background noise and the TEN test. Participants were compensated for their time on a per session basis.

1. Sentence recognition testing

Sentence recognition was assessed using the connected speech test (CST; Cox *et al.*, 1987, 1988). The CST uses everyday connected speech as the test material and consists of 28 pairs of passages (24 test and 4 practice pairs). The recommended key word method of scoring was used. Each passage pair contained 50 key words. A total of two passage pairs were completed for each condition, and the score for each condition was based on the average result of these two passage pairs (i.e., based on 100 key words).

Sentence recognition was assessed at multiple low- and high-pass filter cutoff frequencies (total of 10–12 filter conditions) in order to obtain performance versus filter cutoff frequency functions. These functions were used in the derivation of crossover frequency for each subject. Crossover frequencies (defined as the filter cutoff frequency at which the score for low- and high-pass filtered speech is the same) allow for the comparison of the relative importance of low- and high-frequency information between groups and thus directly test whether hearing loss results in a frequency-specific deficit in the contribution of speech information.

In all low-pass filter conditions the high-pass filter cutoff frequency was fixed at 178 Hz. Likewise, in all high-pass filter conditions the low-pass filter cutoff frequency was fixed at 6300 Hz. All subject groups completed the following six filter conditions: low pass 1600 and 3150 Hz, high pass 1600 and 800 Hz, wideband (178–6300 Hz), and bandpass (800–3150 Hz). Performance was assessed on each subject in an additional four/five filter conditions to provide a better indication of the performance versus filter cutoff frequency function for each subject. Specifically, speech recognition of the NHU group was also assessed at low-pass filter cutoffs of 800 and 1200 Hz, as well as high-pass filter cutoffs of 2000 and 2400 Hz. The NHS group completed additional speech recognition testing at low-pass filter cutoffs of 800 and 2000 Hz and high-pass filter cutoffs of 2000 and 3150 Hz. The precise filter cutoff frequencies and the total number of filter conditions tested varied somewhat between HI participants due to differences in absolute performance levels. That is, additional cutoff frequencies were inserted in order to obtain the most accurate estimate of crossover frequency. Each HI subject completed a total of two to five additional filter conditions. All testing was completed in two test sessions with at least one CST passage completed in each filter condition during a session. This methodology allowed for results between groups to be examined not only in terms of crossover frequency but also based on absolute changes in performance as filter cutoff frequencies changed.

TABLE II. Auditory thresholds (in dB HL) for participants with hearing loss.

		Frequency (in Hz)							
Subject	Ear	250	500	1000	2000	3000	4000	6000	8000
DJ	L	60	60	65	60	65	65	65	85
JA	L	50	60	60	65	65	65	70	65
MB	L	55	55	65	70	65	75	75	75
DG	L	55	60	60	65	65	60	65	60
JM	L	55	60	55	60	60	65	80	80
SG	L	60	65	65	65	60	65	80	90
AL	R	70	65	65	70	65	70	75	65
VP	L	50	55	70	70	65	65	70	70
SP	L	55	60	65	65	65	60	60	70

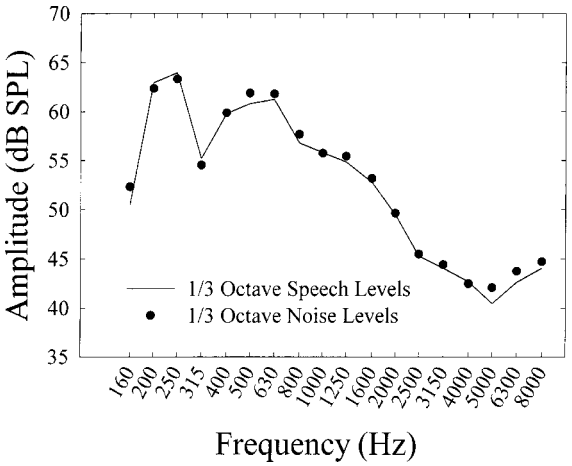


FIG. 1. One-third-octave rms levels (160–8000 Hz) of CST stimuli and spectrally matched noise. The overall levels of the speech and noise were 70 dB SPL.

The masking noise used in this study was created from the calibration noise (track 80, left channel) provided with the audio version of the CST (Hearing Aid Research Laboratory, Speech Intelligibility Tests, 1994). This calibration noise was modified slightly using commercially available sound editing software (Sonic Foundry: SoundForge V. 4.5) to improve the match at 160 Hz. Figure 1 shows the $\frac{1}{3}$-octave rms levels (160–6300 Hz) of the CST talker (based on the entire corpus of CST test passages, tracks 32–55) and the masking noise used in this study.

The speech and noise stimuli were digitally recorded (24.414 kHz sampling rate) and stored on a computer hard disk. Noise was on only during the presentation of speech stimuli (~250–500 ms prior to and following the speech). Digital filtering, employing steep filter skirts (i.e., 800 dB/octave), was used to create the filtered speech and noise stimuli. The Tucker-Davis Technologies (TDT) RP2 Real-Time processor was used for mixing, shaping, and real-time filtering of the speech and noise stimuli. The stimuli were digitally mixed at a +6 dB SNR. The +6 dB SNR level was employed in an attempt to limit ceiling effects among the participants with normal hearing while not causing floor effects for the participants with hearing loss.

To improve audibility for the participants with hearing loss, spectral shaping of the mixed speech and noise stimuli was performed. Subjects in the NHS group also listened to this shaped speech. The spectral shaping was applied to approximate desired sensation level targets (DSL v4.1 software, 1996) for conversational speech, assuming linear amplification for a subject with a flat 65 dB HL hearing loss (Cornelisse *et al.*, 1995). Shaping was verified by measuring the $\frac{1}{3}$-octave rms levels of the masking noise (which spectrally matched the speech) in a Zwislocki coupler using a Larson-Davis 814 sound level meter (slow averaging, C-weighting). The rms value of the difference between coupler outputs and DSL targets (250–6000 Hz) was 1.7 dB, with a maximum difference of −3 dB. The SNR, shaping, and presentation levels of the speech and noise were chosen, in part, to ensure that noise levels rather than quiet thresholds were the primary factor determining audibility for both the HI and NH groups.

The mixed and shaped stimuli were filtered as needed for the appropriate condition prior to being output from the RP2 (24.414 kHz sampling rate). Following output, the filtered speech and noise were low-pass filtered (10 kHz), amplified using a Crown D75 2-channel amplifier, and routed to an ER3 insert earphone (Etymotic Research). Levels were calibrated in the wideband condition (178–6300 Hz) in a Zwislocki coupler and no corrections for level were applied in the various filter conditions. Output levels were measured in a Zwislocki coupler using a Larson-Davis 814 sound level meter (C-weighting, slow averaging).

Output level for the wideband speech was 95 dB SPL for the NHU and NHS groups. The 95 dB level was chosen in an attempt to present speech at levels comparable to those experienced by persons with hearing loss when wearing hearing aids, while at the same time limiting loudness discomfort for the participants with normal hearing. All participants in this experiment reported that the 95 dB SPL presentation level was loud, but not uncomfortable. Wideband levels for HI participants varied slightly depending on individual loudness comfort levels. Seven participants preferred an overall speech level of 103 dB SPL while two participants preferred a higher speech level of 108 dB SPL. Coherence measurements at the output of the earphones indicated minimal distortion (average coherence ~0.95 over the frequencies of 178–6300 Hz) when the speech-shaped noise was presented at the highest speech level used in this study (108 dB SPL).

2. Threshold assessment in noise

Monaural masked thresholds, using the same ear used for speech testing, were determined for the octave frequencies of 250–4000 Hz, as well as the interoctave frequencies of 3000 and 6000 Hz. Masked thresholds in noise were obtained for two primary reasons. First, measuring thresholds in the speech noise helps confirm that the masking noise, rather than quiet thresholds, determined the audibility of speech. Additionally, previous research has shown that masked thresholds of HI persons may be greater than those of NH subjects listening in the same noise. In other words the ''effective masking spectrum'' of the noise may be greater for HI persons than NH persons (i.e., Dubno and Ahlstrom, 1995). Therefore, a second reason for measuring thresholds in the speech noise is to determine the ''effective masking spectrum'' of this noise for each group. This information may be useful in explaining residual differences in speech understanding between groups.

A clinical method of threshold assessment (ASHA, 1978) was used, however, step sizes were reduced (4 dB down–2 dB up) to improve threshold accuracy. The background noise used during threshold testing was the same noise, previously described, used during speech testing and was on continually during threshold assessment. An attempt was made to measure masked thresholds at the same overall masker level as was used during speech testing. Levels, however, varied slightly depending on the loudness tolerance of individual participants. Overall levels of the masking noise,

measured in a Zwislocki coupler using a Larson-Davis 814 sound level meter (C-weighting, slow averaging), were 86–89 dB SPL for the NHS participants and ranged from 94 to 102 dB SPL for the HI participants.

3. Diagnosis of dead regions

Cochlear integrity was assessed for the HI and NHS groups using the CD version of the ''TEN test'' (Moore *et al.*, 2000). Pure-tone thresholds for each subject with hearing loss were measured monaurally (same ear as speech testing) in quiet and in the presence of the ''threshold equalizing noise,'' at octave frequencies of 250–4000 Hz and interoctave frequencies of 3000 and 6000 Hz. In addition, 5000 Hz was also tested for the HI participants. TEN levels varied, depending on loudness discomfort issues, from 65 to 80 dB/ERB (overall levels of 81.5–96.5 dB SPL in a Zwislocki coupler) for the NHS participants and between 75–85 dB/ERB (overall levels of 91.5–101.5 dB SPL) for the HI participants. A subject is considered to have a dead region at a specific frequency if (1) the masked threshold is 10 dB or more above absolute threshold in quiet and (2) the masked threshold is at least 10 dB or more above the noise level per ERB (Moore, 2001).

III. RESULTS AND ANALYSIS

A. Test session effects

Recall that speech data for each participant were collected during two test sessions. To examine potential practice effects, an initial mixed model analysis of variance (ANOVA) was performed. An alpha level of 0.05 was used for this, and all other, statistical analyses reported in this paper. Prior to statistical analysis, individual percent correct scores were converted to rationalized arcsine transform units (RAUs) to stabilize the error variance (Studebaker, 1985). Only the filter conditions that all three groups (NHU, NHS, and HI) completed were used in the analysis (low pass 1600 and 3150 Hz, high pass 1600 and 800 Hz, wideband 178–6300 Hz, and bandpass 800–3150 Hz).

ANOVA results showed a significant between-subjects main effect of group ($F_{1,24}=1097.41$, $p<0.001$) and a significant within-subjects interaction between filter condition and group ($F_{10,120}=2.28$, $p=0.018$). Follow-up testing was not performed at this time since the primary focus of this ANOVA was on differences between test sessions. A significant within-subject main effect of test session ($F_{1,24}=5.69$, $p=0.025$) was also observed. Overall performance was slightly better during the second session (~2.7% improvement), however, no significant interaction effects were observed. Therefore data between test days were collapsed for further analysis.

B. Derivation and examination of crossover frequencies

To determine crossover frequencies, each participant's average sentence recognition scores (in proportion correct) were plotted as a function of the log of their filter cutoff frequency. Then nonlinear regression, using a three-parameter sigmoid function (SPSS, Inc.: SigmaPlot V. 5.0), was used to provide a best fit to the data. Equation (1) shows the function used to fit the data.

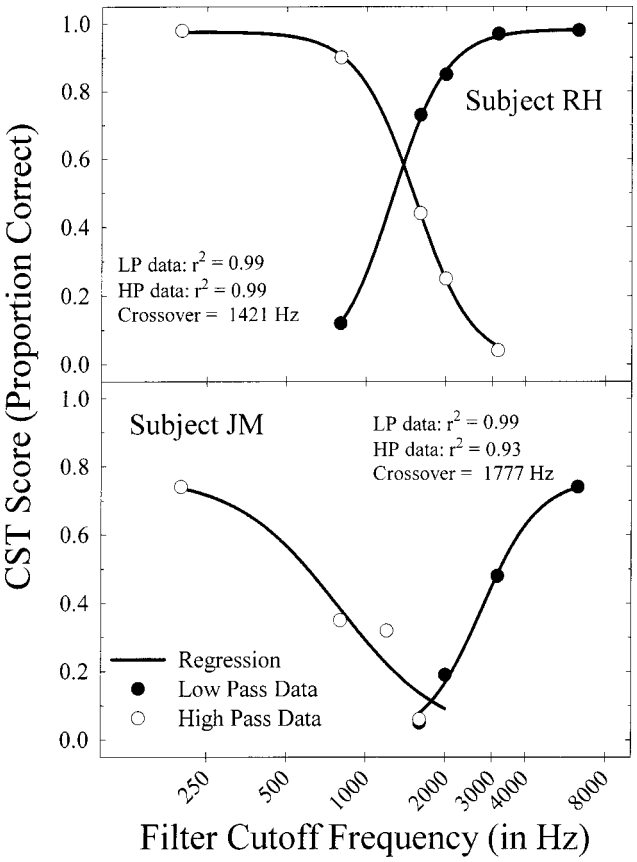


FIG. 2. Examples of how crossover frequencies were determined for both NH (RH) and HI (JM) participants. Nonlinear regression, using a three-parameter sigmoid function (SigmaPlot V5.0) was used to provide a best fit to the low- and high-pass data. Regression functions were then used to determine the crossover point.

$$y=\frac{a}{1+e^{-(x-x0)/b}}. \tag{1}$$

In this formula, y represents the predicted CST score, x represents the filter cutoff frequency and a, b, and $x0$ are free parameters. Figure 2 shows typical single subject results from the NHS (subject RH) and HI (subject JM) groups. The symbols represent the measured scores while the solid lines represent the regression function. The fitted functions were used to determine the crossover frequency.

The polynomial functions for the NH participants provided a good fit to the data, with r^2 values ranging from 0.96 to 0.99. The average scores at the crossover frequency were 40% and 41% for the NHU and NHS groups, respectively. The sigmoid function also provided a good fit to the HI data, although results were more variable with r^2 values ranging from 0.93 (subject JM high pass data) to 0.99. The average score at crossover frequency (14%) was lower for the HI group than for the NHU or NHS groups.

NHU group data was obtained primarily to verify that spectral shaping, appropriate for a person with a flat hearing loss, would not affect the relative importance of different frequency regions to speech understanding. To examine the

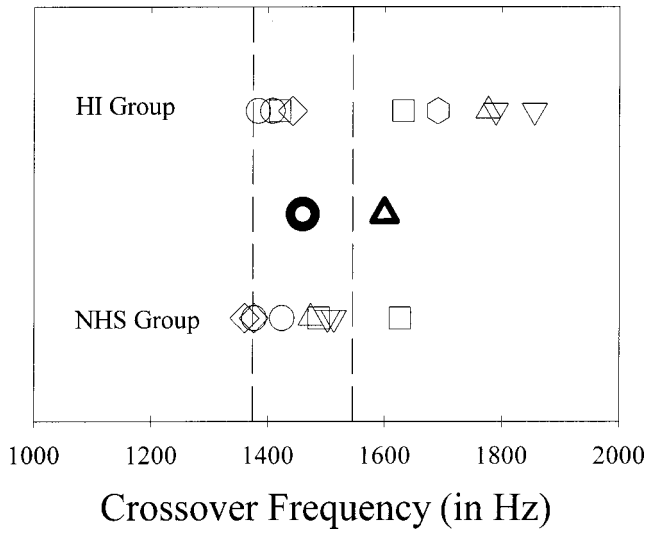


FIG. 3. Crossover frequency data for NH and HI subjects. The dark circle and triangle represent the average crossover frequency for the NH and HI groups, respectively. The remaining symbols represent individual crossover frequencies for the NHS (lower) and HI (upper) groups. Dashed lines show plus of minus 1 standard deviation around the NH mean crossover frequency.

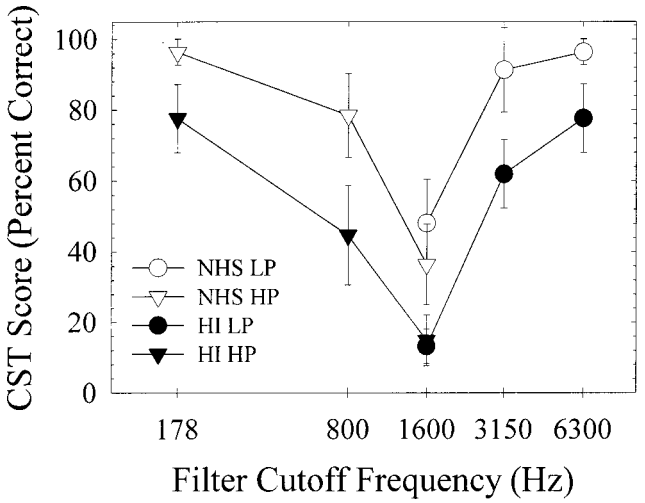


FIG. 4. Average CST scores for NH and HI participants as a function of low- and high-pass filter cutoff frequency. The filled and unfilled symbols show scores for the HI and NH participants, respectively. The circles and triangles represent scores for low- and high-pass filtering, respectively. Error bars show 1 standard deviation.

impact of spectral shaping, crossover frequencies were derived for each subject in the NHU and NHS groups and a single factor between-subjects ANOVA was used to examine differences in crossover frequencies between groups. The independent and dependent variables were subject group and crossover frequency (in Hz), respectively. ANOVA results showed no statistically significant ($F_{1,16}=0.59$, $p=0.45$) differences in crossover frequencies between the NHS and NHU groups. Consequently, all further analysis focuses on results from the NHS and HI groups.

Of primary interest in this study was the effect of hearing loss on the importance of high- and low-frequency information to speech understanding. Recall that one viewpoint suggests that participants with hearing losses are less able to make use of high- versus low-frequency speech information when compared to participants with normal hearing listening under comparable conditions (i.e., similar SNRs and presentation levels). This difference would result in a systematic lowering in crossover frequencies for participants with hearing loss. In contrast, if a uniform deficit occurs across all frequencies, then absolute performance levels of the HI group may be lower, but crossover frequencies should be the same between groups. To examine the effect of hearing loss on crossover frequency, a single-factor between-subjects ANOVA was used to examine differences in crossover frequencies between the HI and NHS groups. The independent and dependent variables were subject group and crossover frequency (in Hz), respectively. Figure 3 shows the mean crossover frequencies for the NHS and HI groups as well as individual crossover frequencies for each HI and NHS participant.

The average crossover frequencies in this study were somewhat higher for the HI group, with mean crossover frequencies of 1600 and 1460 Hz for the HI and NHS groups, respectively. This pattern approached, but did not reach, statistical significance ($F_{1,16}=4.13$, $p=0.059$). The results in Fig. 3 also suggest that the distribution of crossover frequencies for HI participants was bimodal in nature. Specifically, crossover frequencies for the HI appeared to cluster either near the NHS mean (four listeners) or cluster well above the mean (five listeners).

C. Bandwidth effects on sentence recognition

To examine the effect of hearing loss on the contribution of more specific frequency regions to speech understanding, CST performance was compared between the NHS and HI groups as additional low- and high-frequency information was made available (via systematic increases in low- or high-pass filter cutoff frequency). Again, only those filter conditions that participants in both groups completed were used in this analysis. These filter conditions were categorized into three groups for analysis: low pass (LP 1600 Hz, LP 3150 Hz, and broadband), high pass (HP 1600 Hz, HP 800 Hz, and broadband) and band widening (BP 800–3150 Hz, LP 3150 Hz, and HP 800 Hz). Scores (in percent correct) for the low- and high-pass filter conditions are shown in Fig. 4.

A series of planned comparisons, using single factor mixed model ANOVAs, were performed to compare performance differences between groups. The between-subjects independent variable was group (NHS and HI) while the within-subjects independent variable was filter condition. The dependent variable was the participant's average CST score (in RAUs) for each filter condition. Separate ANOVAs were performed on the data in the low-pass, high-pass, and band widening conditions. Summary results are shown in Table III.

A significant between-subjects main effect of group (HI and NHS) was observed in both the low- and high-pass analyses conditions. Average sentence recognition was significantly poorer for the HI group than for the NHS group. Likewise, a significant within-subjects main effect of filter condition was also observed in both the low- and high-pass conditions, with performance improving significantly as the

TABLE III. Results from a series of single factor mixed-model ANOVAs. *P*-levels less than 0.05 are shown in boldface. (LP condition includes filter cutoff frequencies of 1600, 3150, and broadband; HP condition includes filter cutoff frequencies of 1600, 800, and broadband; and BP condition includes 800–3150 Hz, LP 3160, and HP 800; group refers to the normal-hearing shaped and hearing-impaired groups).

Effect	*df*	*F*	*p*-level
	Low pass data		
LP condition	2,32	359.57	**<0.001**
Group	1,16	73.18	**<0.001**
LP×Group	2,32	1.57	0.223
	High pass data		
HP condition	2,32	700.52	**<0.001**
Group	1,16	34.87	**<0.001**
HP×Group	2,32	4.12	**0.026**
	Band pass data		
BP condition	2,32	104.68	**<0.001**
Group	1,16	48.06	**<0.001**
BP×Group	2,32	0.043	0.958

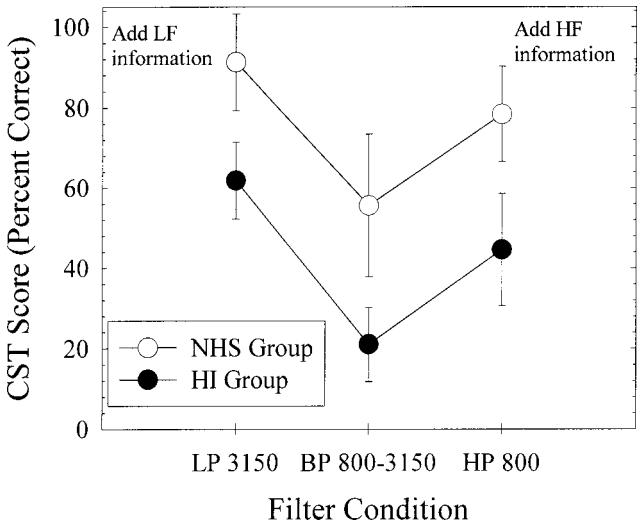


FIG. 5. Average CST scores for NH and HI participants as a function of increasing low- or high-frequency bandwidth. The filled and unfilled symbols show scores for the HI and NH participants, respectively. Error bars show 1 standard deviation.

low- or high-frequency bandwidth increased. Finally, a significant interaction between filter condition and group was also observed for the HP data only. This finding suggests that as the HP filter cutoff frequency was lowered, changes in CST performance were different between the NHS and HI groups. No such interaction was noted in the LP condition, suggesting that on average, the HI and NHS groups made similar use of increases in high-frequency information as the low-pass filter cutoff frequency increased.

Follow-up analysis was performed, using a series of planned comparisons, to determine (1) if the improvements in CST score with bandwidth increases were statistically significant and (2) to determine the cause of the significant interaction in the HP condition. Results revealed that CST scores increased significantly ($p<0.001$) with each increase in filter cutoff frequency regardless of the condition (e.g., LP or HP conditions). Likewise, a significant effect of group ($p<0.001$) was also observed in each follow-up analysis with the HI group performing significantly poorer than the NHS group in each condition. A significant interaction between group and filter condition, however, was present only in the analysis of the HP data when comparing performance in the HP1600 Hz and HP800 Hz conditions ($F_{1,16}=9.76$, $p=0.007$). A review of Fig. 4 shows that in these conditions average scores for the NHS group increased 42% as the high-pass filter cutoff frequency was lowered from 1600 to 800 Hz. CST scores for the HI group, however, increased only 29.8%, resulting in a significant interaction effect between filter condition and group.

The results shown in Fig. 4 also suggest that performance improvements for the NHS group may be limited in some conditions by ceiling effects (i.e., going from HP800 or LP 3150 to broadband). In an attempt to limit the potential impact of ceiling effects, performance was also assessed in a mid-frequency band (BP 800–3150 Hz) condition. The band widening (BW) conditions examined changes in CST scores between groups as additional segments of both low-frequency (178–800 Hz) and high-frequency information (3150–6300 Hz) were added to this mid-frequency (BP 800–3150 Hz) band. Since the wider bandwidth (178–3150 and 800–6300 Hz) used in these comparisons were narrower than the broadband condition (178–6300 Hz), maximum performance levels were also reduced, thus reducing the impact of ceiling effects. Results (in percent correct) are shown in Fig. 5.

Consistent with performance in the LP and HP conditions, significant within-subjects effects of filter condition and significant between-subjects effects of group were observed (see Table III). Overall performance was reduced for the HI, compared to the NHS group, and performance improved for both groups as bandwidth increased, compared to performance listening to the mid-frequency band (800–3150 Hz) alone. Also consistent with performance in the LP conditions, no significant interaction effect between group and filter condition was observed, suggesting that the HI and NHS groups made similar use of the additional high- and low-frequency information, when added to a mid-frequency band of speech.

D. Thresholds in noise

1. Thresholds in speech noise

Masked thresholds were first compared to thresholds in quiet to confirm that the masking noise, rather than quiet thresholds, determined audibility differences between groups. Results showed that quiet thresholds for both the NH and HI groups were shifted at least 5 dB by the masking noise at all frequencies tested with the exception of 6000 Hz. At 6000 Hz the level of masking noise was not intense enough to cause a 5-dB shift in quiet thresholds for six of the nine HI participants. For these HI participants, auditory threshold at 6000 Hz rather than the masking noise dictated audibility in noise.

In addition, differences in the ''effective masking spectrum'' of the background noise, between the NHS and HI groups, were evaluated by subtracting the $\frac{1}{3}$-octave band rms level (dB SPL) of the noise from the level of the pure tone at

TABLE IV. Average SNRs at threshold as a function of frequency for the NHS and HI groups, as well as individual SNRs for participants with hearing loss.

	Average NHS group (in Hz)					
	250	500	1000	2000	3000	4000
NHS	−3.8	−6.3	−6.7	−8.5	−4.6	−6.4
	HI participants					
DJ	1.1	−1.5	1.7	−1.4	3.2	0.8
MB	−0.9	−1.5	−2.3	−1.4	1.2	2.8
DG	−0.9	−3.5	−2.3	−1.4	−2.8	−3.2
AL	9.1	−0.5	2.7	0.6	4.2	1.8
SP	−1.9	−5.5	−1.3	−1.4	0.2	−2.2
JA	1.1	−1.5	−0.3	−3.4	−0.8	−1.2
JM	0.1	−2.5	0.7	−0.4	0.2	−2.2
SG	−1.9	1.5	−3.3	−1.4	−1.8	−5.2
VP	−2.9	−3.5	3.7	0.6	−0.8	−4.2
Average	0.3	−2.1	−0.1	−1.1	0.3	−1.4

threshold (as measured in a Zwislocki coupler). This process provided the SNR required for threshold detection in noise. The SNRs were then compared between groups. Described in this fashion a more negative SNR is better (i.e., lower signal SPL is required for threshold in noise). The average results for the NHS and HI group, as well as individual results for HI participants, are shown in Table IV for the frequencies of 250–4000 Hz.

As seen in Table IV, average SNRs for participants with hearing loss were substantially elevated compared to the NHS group. A mixed model analysis of variance (ANOVA) was performed to examine differences in SNRs between the NHS and HI groups. The within- and between-subjects independent variables for this analysis were frequency (250, 500, 1000, 2000, 3000, and 4000 Hz) and group (NHS and HI), respectively. The dependent variable was the SNR for each participant. A significant main effect of group ($F_{1,16} = 64.201$, $p < 0.001$) was observed with the average SNR for HI participants about 5.4 dB poorer (higher) than for the NHS participants. In other words the "effective masking spectrum" of the speech noise was significantly greater for the HI participants than the NHS participants. This suggests that despite listening at the same acoustic SNR, audibility between groups was not equated. In addition, differences between HI and NHS thresholds did vary as a function of frequency. The group-by-frequency interaction in the ANOVA analysis showed that these frequency-specific differences in SNRs approached, but did not reach, statistical significance ($F_{5,80} = 2.12$, $p = 0.072$).

2. Thresholds in TEN

Although any broadband background noise (i.e., the speech noise described above) may be used to identify suspected dead regions, the TEN test (Moore *et al.*, 2000; Moore, 2001) was specifically designed to allow relatively quick identification and quantification of suspected dead regions. Therefore cochlear integrity was assessed using the TEN. Results for all NHS participants showed that the masking noise produced at least 10 dB of masking and all NHS subject masked thresholds fell within ±5 dB of the expected threshold (based on the level of the TEN). Cochlear integrity in the NHS group appeared to be normal, as expected.

Due to loudness discomfort issues and the severity of hearing loss, results for the HI participants were not as easily interpreted. First, due to loudness discomfort issues, the level of masking noise/ERB could not be raised enough to cause a 10-dB shift in quiet thresholds for many participants. In fact, none of the participants demonstrated at least a 10-dB shift in quiet threshold across all of the eight frequencies tested (250, 500, 1000, 2000, 3000, 4000, 5000, and 6000 Hz), although seven of the nine HI participants did show at least a 10-dB shift at three of the eight frequencies. Of those participants, four showed an abnormal shift in masked thresholds (i.e., thresholds at least 10 dB above the masker level/ERB and 10 dB above threshold in quiet). Results for these four participants are shown in Fig. 6.

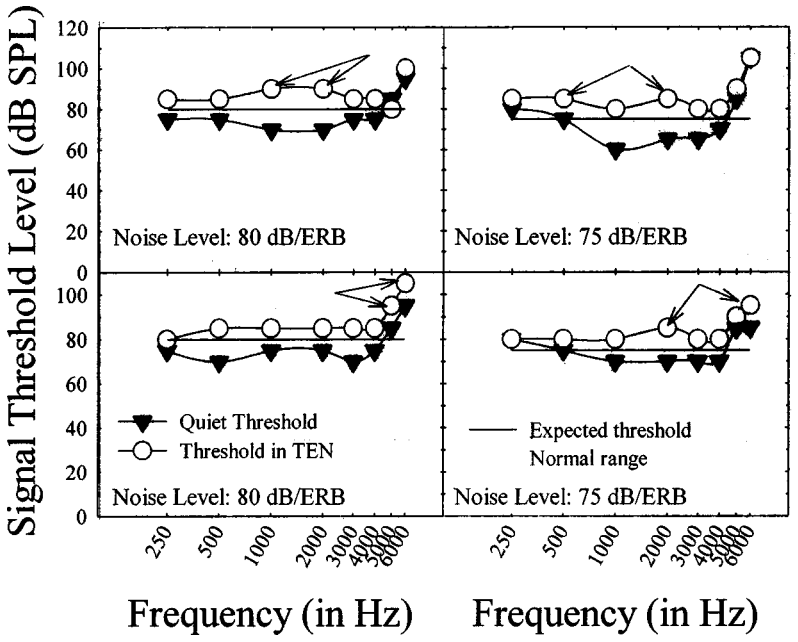


FIG. 6. Thresholds in quiet and in "threshold equalizing noise (TEN)" for four participants with suspected dead regions. Arrows show suspect frequency regions. Triangles and circles show thresholds in quiet and noise, respectively. Solid lines represent expected masked threshold (if masker levels are at least 10 dB above quiet thresholds) and the dashed line represents the normal range of masked thresholds.

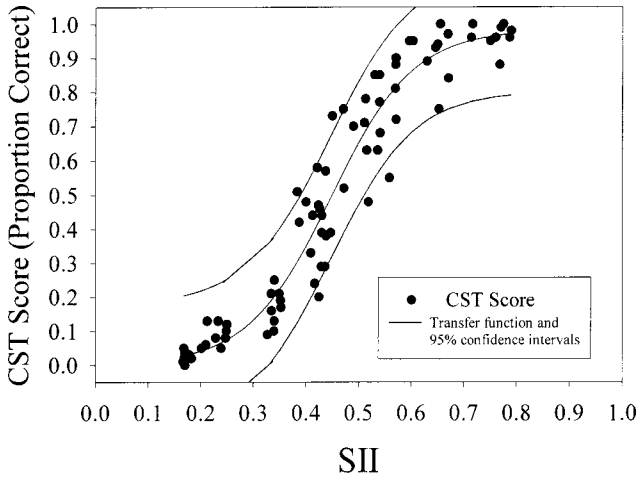


FIG. 7. CST scores as a function of SII for the NHS group. Each data point represents the score of a single NHS subject in one of ten filter conditions. The solid lines show the transfer function and 95% confidence intervals relating the SII to CST performance.

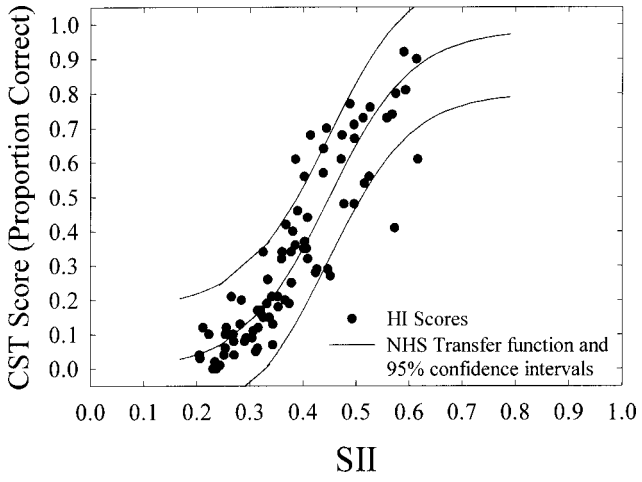


FIG. 8. CST scores as a function of SII for the HI group. The solid lines show the NHS transfer function and 95% confidence intervals from Fig. 7.

E. Speech intelligibility index (SII) calculations

The primary focus of this experiment was to examine whether hearing loss differentially affects the contribution of speech information based on the frequency region of the hearing loss. To make this comparison it is important that audibility, as a function of frequency, be comparable between our NHS and HI groups. Measures of masked thresholds, however, revealed that thresholds were substantially poorer for the HI group compared to the NHS group. In addition, differences in masked thresholds between the NHS and HI groups varied (although not significantly) as a function of frequency. To determine whether differences in audibility between groups could account for the observed performance of the HI subjects, $\frac{1}{3}$-octave band SII calculations (ANSI S3.5, 1997) were performed on the data for each subject. Measures of masked thresholds were interpolated or extrapolated to match the $\frac{1}{3}$-octave center frequencies used in the SII calculations. These thresholds, along with $\frac{1}{3}$-octave rms levels of the speech stimuli, were used to determine the SNR in each $\frac{1}{3}$-octave frequency band. The frequency importance function specifically derived for the CST materials was used (Scherbecoe and Studebaker, 2002). In addition, the effective speech peaks for the CST proposed by Scherbecoe and Studebaker (2002) were used rather than assuming 15-dB peaks as in the ANSI standard. The correction for high presentation levels proposed in the ANSI standard was also included in the calculations, however, corrections for spread of masking effects were not since measures of masked threshold already account for these effects.

1. SII results for the NHS group

CST scores, for the NHS group, as a function of SII are shown in Fig. 7. Each data point represents the score of a single NHS subject in one of the ten filter conditions previously described. The solid lines in Fig. 7 show a transfer function and 95% confidence intervals relating the SII to CST performance.

The transfer function was derived using Eq. (1) with x representing the SII value in a given filter condition for the NHS subjects. The remaining variables a, b, and $x0$ were free parameters that were iterated on to find a best fit to the data. The measure of variance accounted for by Eq. (1) was high with an r^2 value of 0.94.

2. SII results for the HI group

Utilizing the transfer function derived from the NHS data, predictions of individual HI performance were then compared to actual performance across the various filter conditions. Figure 8 shows the individual scores of HI subjects as a function of SII values along with the transfer function for the NHS group. In addition, average results and SII predictions for the HI subjects in the LP, HP, and BW conditions are shown in panels (a) and (b) of Fig. 9.

Together these figures clearly show that, when listening under comparable conditions of audibility, the performance of the HI persons was consistent with that of the NHS group. The poorer performance of the HI group observed earlier in Fig. 4 appears to be due primarily to the increased effectiveness of the masking noise on the HI subjects, thus reducing the audibility of the speech materials. In other words, the utility of speech information appears to be limited primarily by reduced audibility and the reduction is observed uniformly across all frequencies regions.

It is important to note that the accuracy with which the SII predicts the HI data in this study occurs, in part, because the SNRs used in the SII calculations were based on measured thresholds in noise rather than acoustic measures of noise and quiet thresholds. Figure 10 shows the discrepancy between HI performance and SII predictions when the SII is calculated using the $\frac{1}{3}$-octave levels of the speech and noise and individual subject thresholds in quiet. The transfer function and confidence intervals shown in Fig. 10 were derived using the same procedure described in Sect. III E 1, however, SII values were based on acoustic measures of speech and noise levels and quiet thresholds (rather than measured masked thresholds). Since masked thresholds were not used, corrections for spread of masking effects were included in the calculation along with the correction for high presenta-

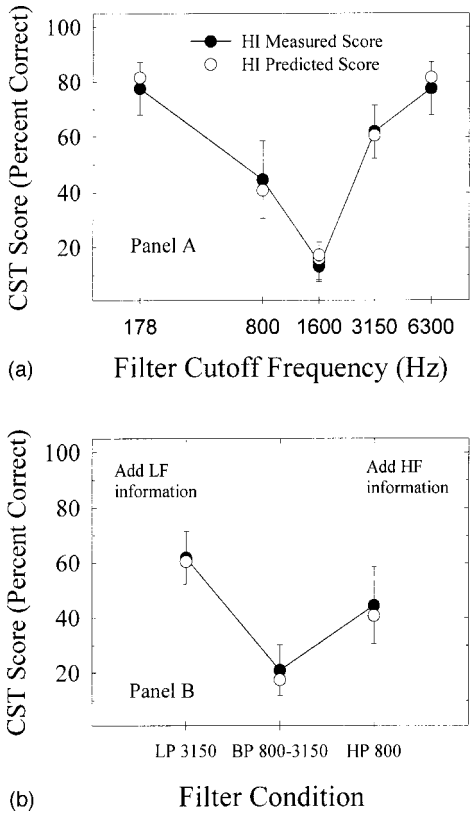


FIG. 9. Panel (a) shows average and predicted CST scores for the HI group as a function of low- or high-pass filter cutoff frequency. Error bars show one standard deviation from the mean measured score in each filter condition. Panel (b) shows the measured and predicted scores from the band widening conditions. The filled and open circles show measured and predicted scores, respectively.

tion levels as described earlier. In this case the SII assumes that acoustic noise levels dictate audibility and that audibility is similar between groups. Therefore the performance of the HI subjects appears to be poorer than predicted based on the NHS transfer function.

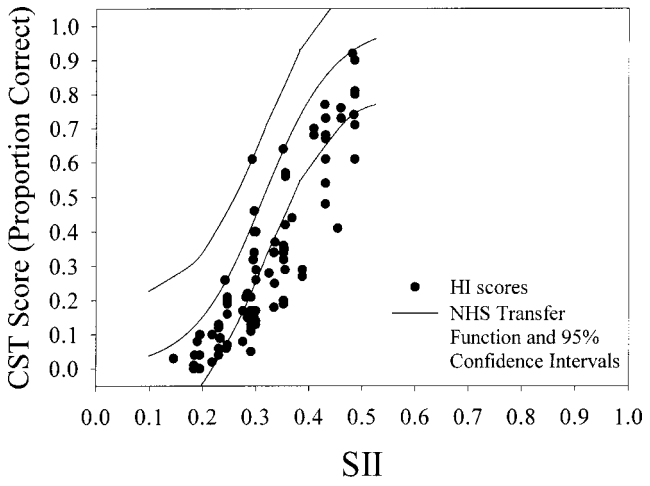


FIG. 10. CST scores as a function of SII for the HI group. The transfer function shown was derived from the NHS group data using Eq. (1) as described earlier. However, SII values for the HI subjects, and for the NHS subjects used to derive the function, were obtained using acoustic measures of the speech and noise levels, as well as each subject's threshold in quiet, to quantify audibility rather than each subject's measured masked thresholds.

IV. DISCUSSION

The focus of this study was on the effects of hearing loss on the importance of various frequency regions to speech understanding. Three potential effects were discussed in the Introduction. Specifically, hearing loss may result in (1) a uniform reduction in the contribution of speech information across all frequencies, (2) a frequency-specific deficit, focused in the high frequencies, and (3) a frequency-specific deficit that varies depending on the presence or absence of cochlear dead regions. The results of this study suggest that, in the absence of dead regions, listeners with relatively flat hearing loss configuration exhibit a more uniform, rather than frequency-specific, deficit in speech understanding.

Support for this conclusion comes from study results showing that, although absolute scores for the HI group were reduced, crossover frequencies for persons with flat hearing losses were comparable to those seen in the normal-hearing control group. The analyses of scores as a function of filter cutoff frequency also add support to the idea that persons with hearing loss are able to make good use of audible high-frequency information to improve speech understanding. On average, CST scores for HI participants improved about 16% when low-pass cutoff frequency was increased from 3150 to 6300 Hz. In contrast, average performance for the NHS group improved only about 5% with the same change in filter cutoff frequency, although ceiling effects played a large role in limiting the NHS improvement. In the band-widening condition, where ceiling effects were less pronounced, improvement in CST scores as the high-frequency bandwidth increased were essentially equivalent for the NHS and HI groups (21% and 24% improvements, respectively). In addition, the results of the SII analysis showed that the NHS and HI groups in this study made comparable use of audible acoustic information across all frequencies tested.

These results are in contrast to some recent work suggesting that moderate to severe hearing loss limits the usability of high-frequency information more than low frequency information (Hogan and Turner, 1998; Turner and Cummings, 1999; Ching *et al.*, 1998, 2001; Amos, 2001). The reason for these findings is unclear; however, several differences between the current research and previous works may have played a role. Such differences include degree of high-frequency hearing loss, configuration of hearing loss (flat versus sloping), and use of a NH control group listening at comparable presentation levels. Other differences between studies that may make comparisons difficult include differences in speech materials, steepness of filter skirts, filter cutoff frequencies, and the presence or absence of noise. The potential impact of some of these factors is discussed below.

One difference between the current study and some previous works, showing a high-frequency deficit for persons with hearing loss, is the degree of high-frequency hearing loss among the participants. In the current study high-frequency hearing loss at 3000–4000 Hz averaged about 65 dB HL and was no greater than 70 dB HL. In contrast, previous studies reporting limited benefit of high-frequency information for persons with hearing loss included some participants with a greater degree of high-frequency hearing loss. For example, participants with flat losses in the Ching

et al. (1998) study had average thresholds at 3000–4000 Hz of about 80 dB HL and participants with severe–profound sloping losses had average thresholds over 100 dB HL. Likewise, some participants with sloping losses in the Hogan and Turner (1998) study had thresholds greater than 70 dB HL at 3000–4000 Hz. It is possible, therefore, that differences in hearing thresholds between studies may have played some role in the discrepant findings. The fact, however, that no HI subjects in this study showed crossover frequencies substantially lower than NH subjects suggests that a cautious approach should be taken when determining a degree of high-frequency hearing loss above which the benefits of amplification may limited.

Another factor distinguishing the current study from the previously cited research is that this study focused on participants with flat rather than sloping hearing losses. It is possible that fundamental differences in the impact of SNHL on speech understanding exist between participants with flat and sloping hearing losses. For example, since participants with high-frequency hearing losses make relatively good use of low-frequency information and speech information is highly redundant across frequencies, persons with high-frequency SNHL simply do not have to rely heavily on the more degraded high-frequency speech cues. This would be especially true when the speech materials were presented in quiet, as in several previous experiments (i.e., Hogan and Turner, 1998; Ching *et al.*, 1998). In contrast, participants with flat losses having some degradation across all frequencies may require a broader range of cues regardless of the frequency region.

Recent work by Turner and Henry (2002) suggest that the presence of background noise, as in this study, may also play an important role in determining the utility of amplified high-frequency speech information. Similar to Hogan and Turner (1998), these authors calculated the ''efficiency'' with which persons with sloping SNHL could make use of increases in high-frequency speech information. Speech understanding was measured at multiple low-pass filter cutoff frequencies in a multitalker babble noise rather than in quiet. In contrast to the Hogan and Turner study, but consistent with the current study, subjects with SNHL listening in noise showed positive efficiencies across all frequency regions regardless of the degree of hearing loss. That is, subjects with high-frequency hearing loss were able to use amplified high-frequency speech information to improve speech understanding regardless of the degree of hearing loss.

The authors suggest that noise substantially limited access to audible speech information and this limitation may explain the discrepant findings between studies completed in quiet and those in noise. When speech is presented in quiet, broad access to ''easy'' speech cues, such as voicing, are available even when the speech is heavily low-pass filtered. As the low-pass filter cutoff frequency is increased the additional information must be used to improve recognition of the more difficult speech features, such as place of articulation. These cues are often difficult for persons with HL even when the speech is presented broadband and in quiet (i.e., Wang *et al.*, 1978). In contrast, when the speech is presented in a noise background, access to easy speech cues may be limited as well. Thus the additional information provided as the filter cutoff frequency is increased can be used to improve recognition of the easier speech cues, resulting in performance improvements in noise that may be reduced in quiet.

Finally, the role that cochlear dead regions play in determining the utility of speech information in various frequency regions remains unclear. Although scores were consistently lower across all filter conditions for the HI participants in this study, results using the TEN test suggested large-scale frequency-specific cochlear damage was unlikely in the participants tested in this study. While the prevalence of dead regions among individuals with SNHL remains unknown, research suggests that high-frequency dead regions are more often associated with severely sloping high-frequency losses or with losses greater than 95 dB in the high frequencies (Moore, 2001; Kapadia *et al.*, 2002). Thus the absence of large-scale high-frequency dead regions among our subjects is not unexpected, given the criteria used for inclusion in the hearing loss group. The results of our work are consistent with that of Vickers *et al.* (2001) and Baer *et al.* (2002) in that our HI subjects did not show evidence of dead regions and were able to make good use of amplified high-frequency speech information. It is possible that subjects in previous studies showing a high-frequency deficit had dead regions in the high frequencies that limited the utility of amplified speech information. The results from the current study, however, show that the presence of cochlear dead regions is not a prerequisite for poorer than normal speech understanding.

V. CONCLUSIONS

In summary, the results of this study lead to the following conclusions. First, high-frequency information (above 3000 Hz) can be quite useful, in terms of speech understanding, for persons with flat SNHL and high-frequency thresholds up to 70 dB HL. Thus a cautious approach is suggested when clinicians are considering limiting gain in the high frequencies, at least for participants with the degree and configuration of hearing loss examined in this study. Second, when hearing loss across frequencies is similar (i.e., for persons with flat HL), the relative importance of each frequency region appears to be similar to that of subjects with NH. Third, SNHL can have a significant negative impact on speech understanding even in the absence of cochlear dead regions. Finally, the overall poorer performance of HI participants in this study is well explained by reduced audibility due to the increased susceptibility to masking seen in these HI participants.

ANSI (**1969**). ANSI S3.5-1969, ''Methods for the calculation of the articulation index'' (American National Standards Institute, New York).
ANSI (**1997**). ANSI S3.5-1997, ''The calculation of the speech intelligibility index'' (American National Standards Institute, New York).
ANSI (**1996**). ANSI S3.6-1996, ''Specifications for audiometers'' (American National Standards Institute, New York).
American Speech-Language-Hearing Association (**1978**). ''Guidelines for manual puretone audiometry,'' ASHA **20**, 297–301.
Amos, N. (**2001**). ''The contribution of high frequencies to speech recognition in sensorineural hearing loss,'' unpublished doctoral dissertation, Indiana University.

Baer, T., Moore, B., and Kluk, K. (**2002**). ''Effects of low pass filtering on the intelligibility of speech in noise for people with and without dead regions at the high frequencies,'' J. Acoust. Soc. Am. **112**, 1133–1144.
Boothroyd, A. (**1967**). ''The discrimination by partially hearing children of frequency distorted speech,'' Inter. Audiol. **6**(2), 136–145.
Boothroyd, A. (**1968**). ''Selection of hearing aids for children,'' unpublished doctoral thesis, University of Manchester, England.
Boothroyd, A. (**1978**). ''Speech perception and sensorineural hearing loss,'' in *Auditory Management of Hearing-Impaired Children*, edited by M. Ross and T. Giolas (University Park, Baltimore), pp. 117–143.
Ching, T. Y., Dillion, H., and Byrne, D. (**1998**). ''Speech Recognition of hearing impaired listeners: Predictions from audibility and the limited role of high-frequency amplification,'' J. Acoust. Soc. Am. **103**, 1128–1140.
Ching, T. Y., Dillion, H., Katsch, R., and Byrne, D. (**2001**). ''Maximising effective audibility in hearing aid fittings,'' Ear Hear. **22**, 212–224.
Cornelisse, L., Seewald, R., and Jamieson, D. (**1995**). ''The input/output formula: A theoretical approach to the fitting of personal amplification devices,'' J. Acoust. Soc. Am. **97**, 1854–1864.
Cox, R., Alexander, G., and Gilmore, C. (**1987**). ''Development of the Connected Speech Test (CST),'' Ear Hear. **8**, 119s–126s.
Cox, R., Alexander, G., Gilmore, C., and Pusakulich, K. (**1988**). ''Use of the Connected Speech Test (CST) with hearing-impaired listeners,'' Ear Hear. **9**, 198–207.
Dubno, J., and Ahlstrom, J. (**1995**). ''Masked thresholds and consonant recognition in low-pass maskers for hearing-impaired and normal-hearing listeners,'' J. Acoust. Soc. Am. **97**, 2430–2441.
Hogan, C., and Turner, C. (**1998**). ''High frequency audibility: Benefits for hearing-impaired listeners,'' J. Acoust. Soc. Am. **104**, 432–441.
Kapadia, S., Blakemore, S., Graumann, D., and Phillips, A. (**2002**). ''Prevalence of cochlear dead regions in patients with high-frequency sensorineural hearing loss,'' Poster presented at the International Hearing Aid Conference, Lake Tahoe, CA, August 2002.
Moore, B. (**2001**). ''Dead regions in the cochlea: Diagnosis, perceptual consequences, and implications for the fitting of hearing aids,'' Trends Ampl. **5**(1), 1–34.
Moore, B., Huss, M., Vickers, D., Glasberg, B., and Alcantara, J. (**2000**). ''A test for the diagnosis of dead regions in the cochlea,'' Br. J. Audiol. **34**, 205–224.
Pavlovic, C. (**1984**). ''Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment,'' J. Acoust. Soc. Am. **75**, 1253–1258.
Pavlovic, C., Studebaker, G., and Scherbecoe, R. (**1986**). ''An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals,'' J. Acoust. Soc. Am. **80**, 50–57.
Scherbecoe, R. and Studebaker, G. (**2002**). ''Audibility-index functions for the connected speech test,'' Ear and Hear. **23**, 385–398.
Studebaker, G. (**1985**). ''A 'rationalized' arcsine transform,'' J. Speech Hear. Res. **28**, 455–462.
Studebaker, G., Scherbecoe, R., McDaniel, D., and Gray, G. (**1997**). ''Age-related changes in monosyllabic word recognition performance when audibility is held constant,'' J. Am. Acad. Audiol. **8**, 150–162.
Turner, C., and Cummings, K. (**1999**). ''Speech audibility for listeners with high-frequency hearing loss,'' Am. J. Audiol. **8**, 47–56.
Turner, C., and Henry, B. (**2002**). ''Benefits of amplification for speech recognition in background noise,'' J. Acoust. Soc. Am. **112**, 1675–1680.
Vickers, D., Moore, B., and Baer, T. (**2001**). ''Effects of low-pass filtering on the intelligibility of speech in quiet for people with and without dead regions at high frequencies,'' J. Acoust. Soc. Am. **110**, 1164–1175.
Wang, M., Reed, C., and Bilger, R. (**1978**). ''A comparison of the effects of filtering and sensorineural hearing loss on patterns of consonant confusions,'' J. Speech Hear. Res. **21**, 5–36.

Wall compliance and violin cavity modes

George Bissinger[a)]
East Carolina University, Greenville, North Carolina 27858

(Received 12 January 2002; revised 14 November 2002; accepted 25 November 2002)

Violin corpus wall compliance, which has a substantial effect on cavity mode frequencies, was added to Shaw's two-degree-of-freedom (2DOF) network model for A0 (''main air'') and A1 (lowest length mode included in ''main wood'') cavity modes. The 2DOF model predicts a $V^{-0.25}$ volume dependence for A0 for rigid violin-shaped cavities, to which a semiempirical compliance correction term, V^{-x_c} (optimization parameter x_c) consistent with cavity acoustical compliance and violin-based scaling was added. Optimizing x_c over A0 and A1 frequencies measured for a Hutchins–Schelleng violin octet yielded $x_c \approx 0.08$. This markedly improved A0 and A1 frequency predictions to within approximately $\pm 10\%$ of experiment over a range of about 4.5:1 in length, 10:1 in *f*-hole area, 3:1 in top plate thickness, and 128:1 in volume. Compliance is a plausible explanation for A1 falling close to the ''main wood'' resonance, not increasingly higher for the larger instruments, which were scaled successively shorter compared to the violin for ergonomic and practical reasons. Similarly incorporating compliance for A2 and A4 (lowest lower-/upper-bout modes, respectively) improves frequency predictions within $\pm 20\%$ over the octet. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1538199]

PACS numbers: 43.75.De [NHF]

I. INTRODUCTION

One confounding factor in making accurate predictions of violin cavity mode frequencies from a physical model has been cavity wall compliance. Generally, wall compliance decreases the mode frequency for the Helmholtz resonance because the pressure acts against the wall ''spring'' lowering the overall cavity+wall stiffness.[1] Numerous experimental and theoretical studies over a long period have supported such an effect for the Helmholtz resonator, and to a greater or lesser extent it would be expected for all the cavity modes. The Rayleigh frequency relationship for the rigid wall Helmholtz resonator, $\omega = c\{A/(LV)\}^{1/2}$ (where A is port area, L is port length, and V is cavity volume), has been around for over 100 years,[2] yet neither this nor later, modified versions explicitly included wall compliance.

A0 has long been considered the most important cavity mode in the violin acoustically. It is often likened to a compliant wall version of the Helmholtz resonance. It has been recognized for quite some time that the Rayleigh relationship often did not predict A0 mode frequencies reliably. John Schelleng, in his landmark 1963 paper ''The violin as a circuit,'' scaled the ''main air'' resonance A0 of the violin to other members of a violin octet[3] using the Rayleigh relationship. Later, serious discrepancies arose between predictions and experiment for octet members, even though the scale was often ''reset'' with close-by instruments.[4]

Two experiments spotlight the failure of the Rayleigh relationship for A0 in a situation where it should have worked best, viz., volume changes *only* through rib height changes: (1) an averaged volume dependence of $V^{-0.33}$ for rib height changes in the four largest violin octet members,[4,5] and (2) a $V^{-0.27}$ dependence for a rigid, violin-shaped cavity with *f* holes.[6] Irrespective of wall compliance, the $V^{-0.5}$ dependence was clearly inaccurate for violin-shaped cavities, and—a telling failure—the rigid cavity result was worse than the compliant cavity.

II. A0–A1 COUPLING

In 1990, Shaw made the first fundamental improvement in the theoretical description of A0 in over a century by including A1, the next higher violin cavity air mode, into a rigid cavity two-degree-of-freedom (2DOF) network theory[7] model. This lowest length mode, with opposite-phase pressure oscillations in the upper and lower bouts, was first characterized in 1973 by Jansson[8] along with many higher violin cavity modes, two of which, A2 and A4, will also be examined here (the second length mode A3 cannot be seen in the violin because the *f* holes create an incompatible nodal region). A1 has a weak inverse dependence of frequency on rib height that later boundary element calculations (closed cavity) supported while introducing a new arching dependence.[9] In 1998 these disparate aspects of A1 were shown to be included naturally in the 2DOF model, as well as—unique among all current models—strong coupling between A0 and A1 through the C-bout inertance.[5]

The 2DOF model treats A0 and A1 as comprising a system with two normal modes, respectively, symmetric or antisymmetric in pressure in the upper and lower bout regions. A0–A1 coupling—analogous to a system of two spring-coupled pendula where changing one member of the system affects the behavior of the other, while still retaining just two normal modes—means that cavity geometry changes directed primarily toward affecting A0 will affect A1, and *vice versa*. In the 2DOF model the presence of A1 partially suppresses the volume dependence[5] of A0 so that $f(\mathrm{A0}) \propto V^{-0.25}$, a much weaker dependence than the (rigid cavity) Rayleigh relationship.

[a)]Electronic mail: bissingerg@mail.ecu.edu

 0001-4966/2003/113(3)/1718/6/$19.00

The successes of Shaw's model make it both the preferred model for A0 and A1, and very difficult to discuss A0 without including A1. As recently as 1963, A1 was dismissed (along with all higher cavity modes) as being ''of little or no value'' acoustically.[3] More recent investigations indicate increasing prominence for A1, partly due to Hutchins retroactively adding it to the ''main wood'' resonance.[4] To date, it has been shown that A1 can (1) radiate well when dropped below all the corpus modes by gas exchange,[10] (2) presumably be related to violin quality[11] via frequency placement relative to a first corpus bending mode (note, however, a fundamental ambiguity here because there are *two* first corpus bending modes), (3) couple strongly to A0, and (4) be a dominant contributor to the acoustic output of a large bass[12] or even a standard violin.[13]

III. VIBROACOUSTIC COUPLING

Coupling between violin cavity modes and the corpus was first reported by Jansson and Sundin[14] in 1974. Cavity-induced corpus motion on the top and back plates for A0 and A1 corresponding to each mode's pressure distribution has been seen in all experimental modal analyses of the violin,[15,16] as well as in this octet, although occasionally the A1 cavity mode coincided with a strong corpus mode and was buried. There is now no question that cavity modes are capable of forcing a significant violin corpus (here taken as top and back plates plus ribs) wall motion, but there is no *a priori* reason to believe that the induced corpus motions will reproduce the pressure distribution. It is, however, an experimental fact that they generally do.

While recognizing that a ''pure'' cavity or corpus mode does not exist for a real violin,[17] and that there is an inherent interaction between the two, experimentally there are violin modes that are clearly classifiable as primarily one or the other. The first corpus bending modes found by Knott[18] in his finite element calculations (*in vacuo*) for a complete violin agreed both in shape and frequency with those observed by Marshall[15] in his pioneering experimental modal analysis of a violin. One can conclude that these are clearly corpus modes, irrespective of any accompanying cavity air motions, with properties dependent on the density, elastic moduli, and shape of the various substructures of the violin as well as the shape of the assembled structure.

Conversely there are cavity modes such as A0 and A1 that force significant wall motion but retain dominant cavity mode characteristics, i.e., their frequency and mode shape depends primarily on interior cavity geometry (including apertures), not violin materials. Significant coupling between cavity modes and the surrounding corpus is expected—and observed. The coupling of structural and acoustic cavity modes depends on the frequency difference as well as the relative spatial congruency between the mode shapes,[19] as was observed for a violin in an experiment employing interior gas exchange to move cavity mode frequencies relative to corpus modes.[20] In highly coupled cases the very concept of ''pure'' corpus or cavity modes must be abandoned.

The vibroacoustic interaction is a difficult problem that is generally handled numerically. Hutchins remarked[4] in 1992 that:

> ''...mathematical parameters based on current measurements and dimensional scaling of violins are not yet adequate enough to do without empirical scaling, especially concerning such factors as wall compliance and cavity frequencies of the violin body,...''

The gap between present theory for rigid cavities and the important practical need to compute reliable compliant-wall cavity mode frequencies is the underlying rationale for this work. By adopting an approach that treats wall compliance like a generalized contributory component to cavity compliance, it is possible to retain the dependence on cavity geometric properties of the rigid-cavity 2DOF model for A0 and A1.

IV. EXPERIMENT AND RESULTS

The interior cavity oscillations of a complete violin octet were analyzed using two small microphones placed in the upper and lower bout regions of each instrument to observe frequency and phase relationships between the upper and lower bout signals. Typically pink or white noise injected via tubing into the upper or lower bouts was used to excite the cavity. Upper/lower bout phase relationships were checked with sine wave excitation at the specific frequencies of A0 and A1. For a few instruments the upper and lower bout measurements gave slightly different (within a few percent) results; the largest separation was for the alto, which had measured values of 267 and 287 Hz. These were averaged for the final values.

Violin cavity mode spectra obtained in a rigid cavity differ significantly from those of an actual violin. Rigid cavity spectra exhibit sharp, well-defined peaks for all expected cavity modes, whereas the actual violin internal measurements show only A0 and A1 clearly. Higher violin cavity modes are interspersed with other spectral structures from corpus-induced cavity oscillations, substantially reduced in frequency, and much less prominent. (A gas-exchange experiment incorrectly labeled A2 and higher modes because of such difficulties[20]). With no model predictions for A2 and A4 frequencies in actual instruments, and with octet instruments ranging so widely in size, no reliable identifications of A2 and A4 could be made from interior cavity measurements, so none were attempted.

There is a reliable indirect way to locate A2 and A4, however. In 1985, Marshall observed not only A0 at 278 Hz and A1 at 478 Hz but also, somewhat surprisingly, A2 at 840 Hz, the A2 value being far below the rigid cavity value of $\sim$1050 Hz.[8] Such a large frequency drop made the identification somewhat suspect, but similar results for other violins and the aforementioned gas-exchange experiments all supported Marshall's original identification. These corpus motions were seen over the entire octet ''mimicking'' A0, A1, A2, and A4 pressure distributions.[12] For the alto the A1 mode was seen only for the 287 Hz (lower bout) A1 cavity value. In Table I the frequencies for A0, A1, A2, and A4 are tabulated for the direct and/or indirect measurements for each octet member, along with f_{low}, the pitch of the lowest string for each octet member.

TABLE I. Measured violin octet A0, A1, A2, and A4 cavity mode frequencies (Hz)—direct measurements with internal microphones, indirect from modal analysis. The pitch, f_{low}, of the lowest string is also shown for reference.

		Direct		Indirect			
Instrument[a]	f_{low}	A0	A1	A0	A1	A2	A4
Treble (sus 123)	391	476	700	not seen	703	1554	1894
Soprano (sus 301)	261	361	560	361	"buried"	930	1295
Mezzo (sus 153)	196	268	422	267	"buried"	629	897
Alto (sus 137)	130	185	277[b]	184	287	414	499
Tenor (sus 151)	98	120	215	121	215	325	459
Baritone (sus 198)	65	91	160	92	154	224	284
Small bass (sus 172)	55	70	128	70	129	177	222
Large bass (sus 178)	41	48	92	48	95	141	166

[a]Hutchins instrument label in parentheses.
[b]Average of 267 (upper bout) and 287 (lower bout).

V. VIOLIN OCTET GEOMETRY

All cavity mode frequency calculations require basic cavity geometry for the computation of mode frequencies. Geometries for a violin octet, along with measured A0 and A1 frequencies (plus A2 and A4), provide the widest practical bowed string instrument size and pitch range. All parameters relevant to Shaw's 2DOF model for A0 and A1 and the rigid cavity semiempirical equations of Bissinger[6] for A2 and A4 are given in Table II. The C-bout inertance length used by Shaw was scaled from the cavity length (other optimized parameters specific to the 2DOF model were taken from Ref. 5). Various semiempirical models require subsets of these parameters. The four largest octet instruments all had rib heights considerably reduced from their original values in attempts to raise A0 to the desired placement.[4]

VI. DISCUSSION

The A0 and A1 modes are now the most studied, best understood, and most interesting acoustically of all the cavity modes. They are important to understanding "violin sound," and key to scaling it to other pitch ranges. A0, the "main air" resonance that dominates the sound of the violin at the lowest pitches, probably will always remain the most important cavity mode—due to its placement below all the corpus modes and its acoustic strength—but it cannot be treated in isolation from A1.

Characterizing wall compliance simply is a difficult task. Averaged wall compliance for any particular cavity mode should be sensitive to wall material properties (elastic moduli and density), arching/curvature, and possibly even the frequency/shape relationship with close-by corpus modes.[19,20] The fact that bowed string instruments typically share similar shapes and similar materials in their various substructures simplifies this matter somewhat. One of the most important applications of cavity mode frequency calculations is in the scaling of the vibratory/acoustic properties of one instrument to another pitch range; some insight into how wall compliance affects these calculations comes from a closer examination of Schelleng's scaling procedure for the octet.

TABLE II. Violin octet cavity geometry parameters. All dimensions are interior, in cm: L = cavity length (disregarding end blocks), W_{ub}/W_{lb} = upper-/lower-bout maximum width, W_{Cb} = C-bout minimum width, H_R = rib height, H_A = average arch height, l = top plate thickness at f holes, V = computed volume (cm^3), A_f = computed total f-hole area (cm^2).

Instrument	L	W_u	W_{Cb}	W_{lb}	H_R[a]	H_A	l	V	A_f
Treble (sus 123)	28.6	12.4	8.0	15.7	1.9	0.93	0.5	800	14.6
Soprano (sus 301)	30.9	14.8	9.6	19.3	1.9	1.00	0.4	1060	12.4
Mezzo (sus 153)	37.8	17.0	11.2	22.4	2.2	1.23	0.35	1760	13.4
Alto (sus 137)	49.7	23.5	15.5	30.0	3.9	1.80	0.4	5410	27.2
Tenor (sus 151)	64.3	30.1	19.9	38.4	5.5	1.93	0.5	12120	31.0
Baritone (sus 198)	85.3	40.0	26.5	52.2	7.7	2.23	0.85	29450	63.0
Small bass (sus 172)	104	48.0	31.0	62.2	9.8	2.83	0.8	54240	72.0
Large bass (sus 178)	128	60.8	38.6	78.2	12.0	3.45	0.9	103 300	112.4

[a]Rib heights for four largest instruments much lower than original values (see Ref. 4).

A. Implications of scaling

Schelleng's network-based scaling of the violin's "main air" and "main wood" resonances rested on two main assumptions: flat plates and "similarity of shape." Schelleng employed simple geometric relationships between the violin's dimensions and those of each member of the octet to place each resonance at the desired frequency, $1.5\times f_{low}$ for the "main air," A0, and $2.25\times f_{low}$ for the "main wood" (which includes A1).

Flat plate scaling of the "main wood" resonance created equations for parameters such as stiffness, plate thickness, mass of corpus, and ultimate strength, based on the ratio of instrument length to that of the violin. Schelleng predicted that wall compliance in the octet increased with instrument size. Similarity of shape means that plate width and rib height scale similar to the length. Hence, scaling suggests an interrelation between wall compliance and instrument volume that effectively associates wall compliance with any geometry-based cavity mode analysis.

B. A0 and A1

1. A0 and A1—rigid walls

A0 and A1 share the whole cavity and 2DOF A0 and A1 frequencies come from solving the same set of equations, so, strictly speaking, logic suggests working with A0 and A1 simultaneously. The original 2DOF predictions for an 11-instrument set that covered a range of 1.5 octaves (versus 3.5 for the octet) agreed surprisingly well with experiment, with $f(\mathrm{A0})$ falling consistently about 11% high, and $f(\mathrm{A1})$ about 15% high,[5] supporting the underlying generality of 2DOF model parameters optimized on a rigid cavity with no arching. In Fig. 1 we compare the ratio of 2DOF frequencies to experiment, 2DOF/exp, for the octet, using the 11-instrument normalization factor of 0.90 for A0 and 0.87 for A1 2DOF predictions. Overall, the agreement between rigid cavity predictions and experiment is not very good. The trends for A0

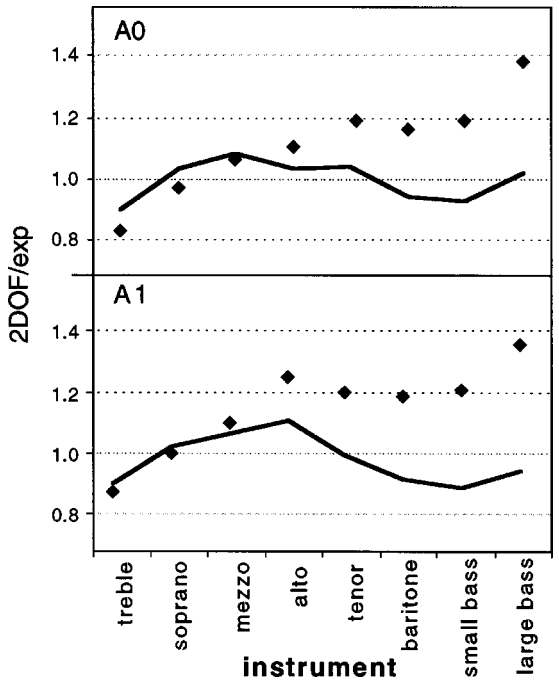


FIG. 1. Top: ratio of 2DOF (♦) and 2DOFC (solid line) predictions to experimental A0 frequencies for the violin octet. Bottom: the same as above but for A1. Typical errors smaller than points.

and A1 are similar, ranging from about 15% low for the smallest to about 37% high for the largest instruments.

2. A0 and A1 with compliant walls

In 1937, Meinel measured violins with, then without, a sound post, and with too thick, then too thin, plates.[21] A downward shift in A0 was noticeable in each case, as it was for more recent experiments such as the removal of physical restraints on the plates[22] or modal analysis of a violin with and without a sound post.[16] The major change in these experiments was clearly the wall compliance. One can reasonably assume similar wall compliance effects for A0 *and* A1 when averaged over the cavity, even though the corpus flexural response is different. Octet A0 and A1 cavity-mode-induced mobilities increased with instrument size relative to the average B1 corpus mode mobility,[12] consistent with the expected trend of compliance increasing with instrument volume.

A1 is a mode whose geometry dependence is predominantly inverse with length.[6] For practical and ergonomic considerations all the larger instruments were made short compared to the ''perfectly scaled'' violin (cf. Fig. 9, Ref. 4). For example, the large bass was only 0.6 of the scaled violin length. Hence, it should have frequencies $\approx 0.6^{-1} \times (2.25 \times f_{\mathrm{low}})$ for the large bass. Yet $f(\mathrm{A1})$ fell close to the desired placement over the entire octet. Wall compliance offers a plausible explanation for this behavior.

3. Incorporating compliance

Network theory employs the mechanical mass and stiffness correlates of inertance (inductance) and acoustical capacitance ($\propto 1/V$) determined directly from cavity geometry to compute cavity mode frequencies. Since wall compliance reduces the effective cavity ''stiffness,'' some form of inverse dependence on volume could also be expected here. Our choice of a functional form for compliance corrections assumes frequencies computed as the product of two functions, i.e., a 2DOF rigid cavity term times a compliance correction term. This bypasses many mathematical and conceptual difficulties encountered in a rigorous compliance–volume link, retains the rigid cavity limit of $V^{-0.25}$, yet maintains a geometry-based approach suitable for cavity modes. Hence the new volume dependence is written as $f \propto V^{-(0.25+x_c)}$, where x_c is the optimized compliance parameter determined from an analysis of our octet experimental results.

4. x_c from rib height changes

A0 frequency increases were achieved earlier on the four largest octet members entirely by cutting back the ribs.[4] Combining the results for all instruments provided a weighted mean value[5] for the overall volume exponent of $0.33 \pm 0.03 = 0.25 + x_c$. Since this overall exponent was determined from frequency/volume *ratios*, all dependences other than volume cancel. Wall compliance increased the volume dependence of $f(\mathrm{A0})$ from $x = 0.25$ (rigid cavity) to $x \approx 0.33$; hence $x_c \approx 0.08$. Hutchins did not collect measurements of $f(\mathrm{A1})$ for the octet instruments, so no equivalent A1 analysis is possible.

5. x_c from optimization

An alternative, independent way to determine x_c from our measurements employs A0 and A1 results. Using the ratio of 2DOF-to-experiment results for A0 and A1 shown in Fig. 1, a normalized range ((max-min)/average) based on this ratio was computed for the octet. Multiplying this ''rigid cavity'' range by a wall compliance term V^{-x_c}, x_c was then varied to search for a minimum. The result was $x_c \approx 0.075$ for A0 and A1, individually or together. This is quite close to the $x_c \approx 0.08$ value estimated from volume change data for the four largest instruments. The 2DOF/exp range computed for A0 was 0.50; with the compliance correction term $V^{-0.08}$, it dropped to 0.18. This simple compliance modification predicted $f(\mathrm{A0})$ values with a maximum error of $\pm 9\%$ (see Fig. 1) for a group of instruments with a range of about 4.5:1 in length, 10:1 in *f*-hole area, 3:1 in top plate thickness, and 128:1 in volume, a marked improvement over prior models, which could not predict the A0 frequency *change* associated with rib height changes in the small or large bass to within 9%. Similar improvements were seen for $f(\mathrm{A1})$, also shown in Fig. 1, with a maximum error of about $\pm 11\%$. Normal variability among instruments still remains a consideration, e.g., our large bass, with $H_R = 12.0$ cm, had $f(\mathrm{A0}) = 48$ Hz, vs 52 Hz for another large bass with $H_R = 12.5$ cm (Table III, Ref. 4). This compliance-modified 2DOF model is labeled 2DOFC.

With proper normalization, formulas for compliant wall 2DOFC frequencies for A0 and A1 are given by (at $T = 20\,^{\circ}\mathrm{C}$; all volumes in cm^3),

$$f_{2\mathrm{DOFC}}(\mathrm{A0}) = 1.68 \times V^{-0.08} f_{2\mathrm{DOF}}(\mathrm{A0}), \tag{1}$$

$$f_{2\mathrm{DOFC}}(\mathrm{A1}) = 1.53 \times V^{-0.08} f_{2\mathrm{DOF}}(\mathrm{A1}). \quad (2)$$

6. A1/A0 ratios

Compliant walls generally reduce cavity mode frequencies. Intra-instrument ratios of frequencies should reduce the effects of wall compliance. Such ratios should be particularly effective for A0 and A1 because they share the entire cavity and an averaged wall compliance. The 2DOFC compliance correction, $V^{-0.08}$, for A0 and A1 obviously cancels out in such a ratio. A comparison of 2DOFC (or 2DOF) ratios with experiment showed agreement to within 4%–10% over the octet.

C. A2 and A4 with compliance

There is relatively little in the musical acoustics literature about A2 or A4 due to their acoustic insignificance for the violin. Jansson characterized A2 and A4 as transverse modes in the lower and upper bouts, respectively.[8] Their mode frequencies should show a primary dependence on the lower or upper bout width, resp., i.e., $f(\mathrm{A2}) \propto n/W_{\mathrm{lb}}$, or $f(\mathrm{A4}) \propto n/W_{\mathrm{ub}}$, where $n = 1,2,3...$. Boundary element calculations[9] agreed quite well with Jansson's results for an encased (rigid wall) violin. Since all prior work on A2 and A4 had been done with rigid cavities,[6,8,9] A2 had not been identified for an actual violin prior to Marshall's initial identification of a violin "corpus" mode at 840 Hz as being induced by A2.[15] Firth's cavity modal analysis inside an actual violin did not extend beyond 800 Hz[23] and thus showed no evidence of A2 or A4. However, subsequent modal analyses on other violins as well as on the octet have seen such induced motions for A2 and A4. A semiempirical analysis based on a water-filled violin cavity experiment had also shown a weak dependence on rib height H_R for each mode. At present, the only equations for calculating $f(\mathrm{A2})$ and $f(\mathrm{A4})$ are semiempirical (SE) ones for rigid cavities,[6] viz.,

$$f(\mathrm{A2}) = \frac{c}{1.514 W_{\mathrm{lb}} + 0.515 H_R}, \quad (3)$$

$$f(\mathrm{A4}) = \frac{c}{1.61 W_{\mathrm{ub}} + 0.273 H_R}, \quad (4)$$

where c is the velocity of sound in air.

The wall compliance effect on these mode frequencies for the violin can be estimated by comparing the average A2 and A4 frequencies from prior modal analyses, $f(\mathrm{A2}) = 830 \pm 29$ Hz, and $f(\mathrm{A4}) = 1067 \pm 32$ Hz, with rigid cavity values of ~1050 and ~1290 Hz, respectively.[8] Both A2 and A4 share a similar normalization factor of 0.81 in Eqs. (3) and (4) for a comparison with the actual instruments. Compliant walls appear to reduce the violins' A2 and A4 frequencies on the order of 20%. However, compliance increases with size so that larger instruments should have larger relative decreases for $f(\mathrm{A2})$ and $f(\mathrm{A4})$.

Figure 2 shows ratios, SE/exp, of semiempirical A2 and A4 frequencies [from Eqs. (3) and (4) using the normalization factor 0.81] to experimental A2 or A4 results. Again, similar trends appear—predictions relatively low for small—and high for large—instruments. Since the rib height contributions in Eqs. (3) and (4) are relatively small (<5% of the bout-width dependence), $f(\mathrm{A2})/f(\mathrm{A4}) \approx W_{\mathrm{ub}}/W_{\mathrm{lb}} \approx 0.8$, a result quite close to experimental results for the rigid violin cavity, the violin, and all octet members. This broad agreement, combined with the very weak rib height dependence in Eqs. (3) and (4), argues for dropping the rib height terms in Eqs. (3) and (4), with essentially no loss of overall accuracy in the compliance-corrected equations.

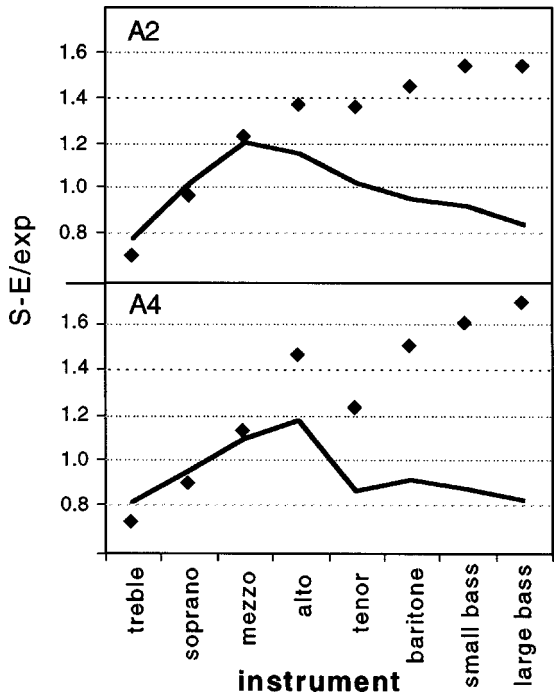


FIG. 2. Top: ratio of SE [Eq. (3)×0.81, ◆] and SEC [Eq. (5), solid line] model predictions to experimental A2 frequencies for the violin octet. Bottom: the same as above but for A4 [and Eqs. (4) and (6), resp.]. Typical errors are smaller than points.

Although A2 and A4 induce similar across-the-bout flexure, it is different in character than for A0 and A1. Hence, a new normalized range calculation was performed for the A2 and A4 datasets simultaneously. The resulting compliance correction was $x_c = 0.16$, resulting in compliance-corrected semiempirical (SEC) equations (c in centimeters per second, all dimensions in centimeters),

$$f(\mathrm{A2}) = \frac{1.68c}{V^{0.16} W_{\mathrm{lb}}}, \quad (5)$$

$$f(\mathrm{A4}) = \frac{1.68c}{V^{0.16} W_{\mathrm{ub}}}. \quad (6)$$

These equations offer considerably improved agreement with experiment, as can be seen from Fig. 2, although not quite as good as for A0 and A1. All compliance-corrected A2 and A4 frequencies now fall within ±20% of experiment and retain a pure geometry dependence in the ratio.

VII. CONCLUSIONS

Elastic walls in a complex shape like the violin cannot be rigorously incorporated into any model of violin cavity modes in a simple way. Instead, wall compliance was added to the rigid-cavity 2DOF model as a semiempirical volume-dependent term $V^{-0.08}$ multiplying the 2DOF A0 and A1

frequencies. This compliance correction is helpful in understanding why earlier equations for the Helmholtz resonance had some success, but failed in situations where they should have worked best. In the 2DOF model A0–A1 coupling changes the volume dependence of $f(\mathrm{A0})$ from $V^{-0.5}$ to $V^{-0.25}$, but wall compliance moves it to $V^{-0.33}$. Scaling A0 for the octet with the Rayleigh relationship, constantly resetting the scaling to the closest instrument as did Schelleng, should give reasonable results. In going from one instrument to another, e.g., violin to cello, where there are additional, substantial *f*-hole, wall thickness, arching, and rib height changes, volume dependence could also be partially subsumed into other dependences. The fundamental inadequacy linked to wall compliance was exposed only when the volume was changed solely through rib height changes, all other dependences essentially canceling in the frequency ratio.

As for A1, for a ''rigid'' large bass it should fall at about $3.75\times f_{\mathrm{low}}$, not the $2.25\times f_{\mathrm{low}}$ observed. If octet members were scaled as a rigid-walled cavities, $f(\mathrm{A1})$ would change from lower-than-desired to higher-than-desired, exactly as seen in Fig. 1. But with wonderful synergism, wall compliance *drops* A1 mode frequencies proportionately more as the instruments grow in size, and thus compensates for the increasingly too-short cavity lengths of the larger instruments, the net result being that A1 always falls close to the ''main wood'' resonance, as in the violin. Of course, a small contribution from the aforementioned $f(\mathrm{A1})$-lowering effect of relatively lower arches for the larger instruments can be expected as well.

At this stage in bowed string instrument development it seems improbable that any practical bowed string instruments will be made larger than the large bass, or smaller than the treble violin (other than those for small children). Within this range it is now possible to place the lower cavity modes in a properly constructed instrument close to a desired frequency, minimizing later cut-and-try trimming. Superior underlying physics, A0–A1 coupling, enhanced sensitivity to details of cavity geometry and wall compliance compensation all combine in the 2DOFC model to predict A0 and A1 frequencies that agree with experiment to within about $\pm 10\%$ across an enormous size range. For higher modes such as A2 and A4 semiempirical relations like Eqs. (5) and (6) are still the best.

ACKNOWLEDGMENTS

I am indebted to Carleen Hutchins for the loan of a complete violin octet for an extended period. We also gratefully acknowledge the support of the National Science Foundation (Grant No. DMR-9802656) for the VIOCADEAS Project of which this research is only a small part. We also wish to acknowledge discussions with Gabriel Weinreich and the help of John Keiffer in the octet EMA data acquisition and Amy Bissinger in the modal data analysis.

[1] L. Cremer, *The Physics of the Violin* (MIT Press, Cambridge, 1984), Chap. 10.
[2] J. W. S. Rayleigh, *The Theory of Sound*, 2nd ed. (Dover, New York, 1945), Vol. 2, Art. 303.
[3] J. Schelleng, ''The violin as a circuit,'' J. Acoust. Soc. Am. **35**, 326–338 (1963); erratum, p. 1291.
[4] C. M. Hutchins, ''A 30-year experiment in the acoustical and musical development of violin-family instruments,'' J. Acoust. Soc. Am. **92**, 639–650 (1992).
[5] G. Bissinger, ''A0 and A1 coupling, arching, rib height, and *f*-hole geometry dependence in the 2-degree-of-freedom network model of violin cavity modes,'' J. Acoust. Soc. Am. **104**, 3608–3615 (1998).
[6] G. Bissinger, ''The effect of violin cavity volume (height) changes on the cavity modes below 2 kHz,'' CAS J. **2**, 18–21 (1992).
[7] E. A. G. Shaw, ''Cavity resonance in the violin: network representation and the effect of damped and undamped rib holes,'' J. Acoust. Soc. Am. **87**, 398–410 (1990).
[8] E. Jansson, ''On higher air modes in the violin,'' Catgut Acoust. Soc. Newsletter, No. 19, 13–15 (1973); E. Jansson, ''Acoustical properties of complex cavities. Prediction and measurements of resonance properties of violin-shaped and guitar-shaped cavities,'' Acustica **37**, 211–221 (1977).
[9] G. Bissinger, ''Acoustic normal modes below 4 kHz for a rigid, closed violin-shaped cavity,'' J. Acoust. Soc. Am. **100**, 1835–1840 (1996).
[10] G. Bissinger and C. M. Hutchins, ''A1 cavity-mode-enhanced fundamental in bowed violin and viola sound,'' CAS J. **1**, 11–13 (1988).
[11] C. M. Hutchins, ''A measurable controlling factor in the tone and playing qualities of violins,'' CAS J. **1**, 10–15 (1989); ''The future of violin research,'' CAS J. **2**, 1–7 (1992).
[12] G. Bissinger, ''Normal mode analysis of violin octet scaling,'' *Proceedings of the International Symposium on Musical Acoustics 2001*, edited by D. Bonsi, D. Gonzalez, and D. Stanzial, 2001, pp. 113–116.
[13] G. Bissinger and J. Keiffer, ''Radiation damping, efficiency, and directivity for violin normal modes below 4 kHz,'' ARLO **4**, 7–12 (2003) (http://ojps.aip.org/ARLO/top.jsp).
[14] E. Jansson and H. Sundin, ''A pilot study on coupling between top plate and air volume vibrations,'' Catgut Acoust. Soc. Newsletter, No. 21, 11–15 (1974).
[15] K. D. Marshall, ''Modal analysis of a violin,'' J. Acoust. Soc. Am. **77**, 695–709 (1985).
[16] G. Bissinger, ''Some mechanical and acoustical consequences of the violin soundpost,'' J. Acoust. Soc. Am. **97**, 3154–3164 (1995).
[17] G. Weinreich, C. Holmes, and M. Mellody, ''Air–wood coupling and the swiss-cheese violin,'' J. Acoust. Soc. Am. **108**, 2389–2402 (2000).
[18] G. A. Knott, ''A modal analysis of the violin using MSC/NASTRAN and PATRAN,'' M.S thesis, Naval Postgraduate School, 1987.
[19] F. Fahy, *Sound and Structural Vibration—Radiation, Transmission and Response* [Academic Press, New York, 1987 (paperback)], pp. 252–256.
[20] G. Bissinger, ''Vibro-acoustics in the violin: a gas-MAC attack on the corpus–cavity interaction,'' *Proceedings of the 15th International Modal Analysis Conference*, Soc. Exp. Mechanics, Bethel, CT, 1997, pp. 699–703.
[21] H. Meinel, ''On the frequency curves of violins,'' Akust. Z. **2**, 22–33 (1937).
[22] C. M. Hutchins, ''A study of the cavity resonances of a violin and their effect on its tone and playing qualities,'' J. Acoust. Soc. Am. **87**, 392–397 (1990).
[23] I. Firth, ''Modal analysis of the air cavity of the violin,'' CAS J. **48**, 17–22 (1987).

Simplified models of flue instruments: Influence of mouth geometry on the sound source

S. Dequand, J. F. H. Willems, M. Leroux, R. Vullings, M. van Weert, C. Thieulot, and A. Hirschberg[a)]
Technische Universiteit Eindhoven, Postbus 513, 5600 MB Eindhoven, The Netherlands

(Received 12 December 2001; revised 27 November 2002; accepted 16 December 2002)

Flue instruments such as the recorder flute and the transverse flute have different mouth geometries and acoustical response. The effect of the mouth geometry is studied by considering the aeroacoustical response of a simple whistle. The labium of a transverse flute has a large edge angle (60°) compared to that of a recorder flute (15°). Furthermore, the ratio W/h of the mouth width W to the jet thickness h can be varied in the transverse flute (lips of the musician) while it is fixed to a value $W/h \approx 4$ in a recorder flute. A systematic experimental study of the steady oscillation behavior has been carried out. Results of acoustical pressure measurements and flow visualization are presented. The sharp edge of the recorder provides a sound source which is rich in harmonics at the cost of stability. The larger angle of the labium of the flute seems to be motivated by a better stability of the oscillations for thick jets but could also be motivated by a reduction of broadband turbulence noise. We propose two simplified sound source models which could be used for sound synthesis: a jet-drive model for $W/h > 2$ and a discrete-vortex model for $W/h < 2$. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1543929]

PACS numbers: 43.75.Np, 43.75.Qr [NHF]

I. INTRODUCTION

In a flue instrument, the sound is produced by the interaction of a jet with a sharp edge (called the labium) placed in the opening (mouth) of a resonator (body of the instrument). The jet is formed by flow separation at the end of a slit (the flue channel) and travels along the mouth of the resonator towards the labium. Coupling between jet oscillations in the mouth and acoustical resonances in the body provides self-sustained oscillations of the jet and maintains sound production.

A large variety of flue instruments are discussed by, among others, Castellengo (1976), Campbell and Greated (1987), and Fletcher and Rossing (1998). In a recent review, Fabre and Hirschberg (2000) discuss the literature on physical modeling of flue instruments. As explained in their paper, there exist no accurate models for the flow in the mouth of such instruments. Formal theories of Crighton (1992), Elder (1992), and Howe (1998) are based on a linear response of very crude models of the flow. They cannot predict the amplitude of the oscillations. We therefore use here modifications of simplified nonlinear models. Next to their passive acoustical behavior, which can be described by means of simple acoustical models (Fletcher and Rossing, 1998; Coltman, 1966; Nederveen, 1998; Wolfe, 2001), flue instruments distinguish themselves also by significantly different mouth geometries. Simple models for the source are found in literature. Coltman (1968, 1969, 1976) developed a simplified model of the sound production which is called the jet-drive model. A similar model was used by Powell (1961) to predict the sound production of an edge-tone system. In this model, the jet flow Q_j is separated at the labium into a flow Q_{in} entering the resonator and a flow Q_{out} leaving the resonator. These flows act as two complementary monopole sound sources ($Q_j = Q_{\text{in}} + Q_{\text{out}}$) (Elder, 1973). This jet-drive model is commonly accepted in the literature (Fletcher and Rossing, 1998). In its original form, the jet-drive model includes a momentum drive associated to the volume injection. Coltman (1980) has shown that this effect is much weaker than predicted by the quasi-steady model of Fletcher and Rossing (1998) and Elder (1973). We will therefore ignore this momentum drive. Fabre *et al.* (1996) and Verge *et al.* (1997a, b), however, demonstrated that vortex shedding at the labium was a key damping mechanism and proposed simplified models for this. When this modified jet-drive model (without momentum drive but with vortex damping) is implemented in a global model of the instrument, one can predict within 1 dB the oscillation amplitudes in a recorderlike instrument (Verge *et al.*, 1997a, b; Fabre *et al.*, 1996) at low Strouhal numbers corresponding to the first hydrodynamic mode of the jet. For higher values of the Strouhal number, the jet-drive concept appears to fail as a result of the breakdown of the jet into discrete vortices. This effect was first reported by Fletcher (1976). In the case of a jet/labium configuration without resonator (edge-tone), Holger *et al.* (1977) proposed to describe the jet flow behavior in terms of a Von Kàrmàn vortex street. The model of Holger (1977) has recently be used with some success by Ségoufin *et al.* (2001). While the idea of using such a discrete-vortex model for a flue instrument at high Strouhal numbers was proposed earlier (Hirschberg, 1995; Fabre and Hirschberg 2000), it was only recently used for a whistle by Meissner (2002) in the case of a very thin jet compared to the mouth width. In the present paper, we investigate the use of the combination of a similar discrete-vortex model and the jet-drive model in order to explain the effect of the ratio between the mouth width W and the jet height h on the flow behavior in a simple whistle.

[a)]Author to whom correspondence should be addressed. Electronic mail: a.hirschberg@tue.nl

 0001-4966/2003/113(3)/1724/12/$19.00

In order to simplify the theory, we consider only the maximum of the dimensionless amplitude which corresponds to a critical Strouhal number at which the sound source is in phase with the acoustical velocity through the mouth. We therefore avoid in the jet-drive model the description of the phase lag due to the jet response.

Furthermore, we describe some experimental results obtained from a study of the effect of the mouth geometry on the oscillation amplitude and its stability. The pulsation amplitudes predicted by our model as a function of the ratio W/h between the mouth width W and the jet thickness h are compared to experimental results. The model is described in Sec. II and the experimental work is presented in Sec. III. This includes some flow visualization.

II. SIMPLE MODELS

In this section, we propose a simple model to predict the pulsation amplitudes in a whistle (or ocarina). This flue instrument can be represented, in a first approximation, by a resonator with a closed volume $V=H^2\times L$ (where H^2 and L are the square cross-section and the length of the cavity, respectively) and a mouth opening of cross-section $S_m=H\times W$ along which a jet is grazing (where W is the width of the mouth opening). The system has a single acoustic mode behavior. The excitation of the resonator by the jet oscillations is represented by either a jet-drive model or a discrete-vortex model (Sec. II C).

A. Single mode model

As explained in a previous study on self-sustained oscillations in a Helmholtz-like resonator (Dequand, 2001, 2001b), the resonator is described as a mass-spring system where the mass represents the incompressible flow in the neck of the resonator (small pipe of length L_m and cross-section area S_m), and the spring represents the volume V of the resonator. This volume is not shallow and plane waves can propagate inside. Such a system could be called a ''stopped pipe.'' By considering the particle displacement ζ in the neck (mouth) of the resonator (defined positive when it is directed into the volume V), Newton's law yields

$$M_a\frac{d^2\xi}{dt^2}+R\frac{d\xi}{dt}+K\xi=\mathcal{F}(t), \tag{1}$$

where $\mathcal{F}(t)$ denotes the aeroacoustical forces acting on the resonator and will be determined in the next paragraph. The mass M_a is related to the effective length L_{eff} of the neck (mouth) of the resonator:

$$M_a=\rho_0 L_{\text{eff}}HW, \tag{2}$$

where ρ_0 is the uniform flow density.

By applying the momentum conservation to the neck, the effective length of the neck is expressed as

$$L_{\text{eff}}=\frac{1}{2\pi f_1\rho_0}\left|\frac{p'}{u'}\right|, \tag{3}$$

where u' and p' are the acoustical perturbations of the velocity and the pressure, respectively. Then, by writing the mass conservation between the resonator and the neck and assuming that only plane waves p^+ and p^- propagate in the resonator [$W\rho_0 c_0 u'=H(p^+-p^-)$ where c_0 is the speed of sound], the effective length L_{eff} becomes

$$L_{\text{eff}}=\frac{W}{H}\frac{c_0}{2\pi f_1}\left|\frac{p^++p^-}{p^+-p^-}\right|=\frac{W}{H}\frac{c_0}{2\pi f_1}\cot\left(\frac{2\pi f_1}{c_0}L\right). \tag{4}$$

The damping coefficient R is related to the quality factor $Q_f=f_1/\Delta f_1$:

$$R=\frac{2\pi f_1 M_a}{Q_f}. \tag{5}$$

The quality factor $Q_f=30$ was determined experimentally from measurement of the 3-dB width Δf_1 of the resonance peak at f_1. By definition, the spring constant K is

$$K=M_a(2\pi f_1)^2. \tag{6}$$

In Eqs. (2)–(6), the frequency f_1 is measured. In the next sections, we will omit the subscript 1 and we will denote the frequency by f. In the present work, we identified f_1 with the whistling frequency at the maximum amplitude of a mode.

B. Losses induced by vortex shedding at the labium

Flow separation occurs at the sharp edge of the labium. We consider here the vortex shedding associated with the quasi-steady flow separation of the acoustical flow through the mouth (Fabre, 1996; Verge *et al.*, 1997a, b).

By assuming that the flow separation of the acoustic flow $Q=(d\xi/dt)HW$ (where HW is the cross-sectional area of the mouth) at the labium results in the formation of a free jet [quasi-steady free-jet model proposed by Ingard and Ising (1967)], the effects of vortices can be represented by a fluctuating pressure Δp_v across the mouth of the instrument:

$$\Delta p_v=-\frac{1}{2}\rho_0\left(\frac{Q}{\alpha_v HW}\right)^2\text{sign}(Q), \tag{7}$$

where α_v is the vena-contracta factor of the jet. We choose the value of $\alpha_v=0.6$ as used by Verge *et al.* (1997a, b). The time-averaged power losses induced by the vortex shedding at the labium is then

$$\langle\mathcal{P}_{\text{losses}}\rangle=\langle\Delta p_v Q\rangle=-\frac{1}{2T}\frac{\rho_0 HW}{\alpha_v^2}\int_0^T\left(\frac{d\xi}{dt}\right)^2\left|\frac{d\xi}{dt}\right|dt. \tag{8}$$

Howe (1975) proposed to use the point vortex model of Brown and Michael (1954) in order to describe the vortex shedding at the labium. Using an asymptotic approximation, he obtains an analytical model which predicts a sound production by this vortex shedding. A great advantage of the model proposed by Howe (1975) is that it could explain the difference in generation of harmonics on the sharpness of the edge of the labium as proposed by Nolle (1983). The classical model of Fletcher and Rossing (1998) does not predict an effect of the edge geometry. The model of Howe (1975) can easily be generalized to be applied to an arbitrary edge angle as long as the vortex remains close to the labium. While self-similar solutions of Pullin (1978) and Peters (1993) are available, we expect that a numerical simulation of the vortex path as a function of time is useful. The numerical study of Dequand (2001, 2001b) indicates significant discrepancies

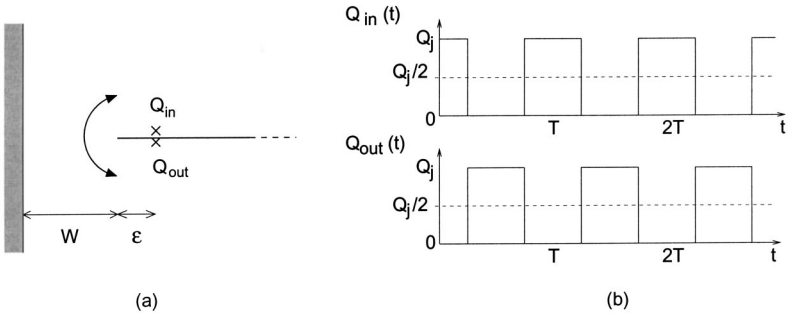


FIG. 1. (a) Idealized representation of the mouth of the instrument in the jet-drive model (Verge *et al.*, 1994a) and (b) time dependence of the two complementary sources Q_{in} and Q_{out}.

between the asymptotic analytical solution of Howe (1975) and a numerical solution. From simple scaling arguments one can, however, expect without calculations that a sharper edge as found in the recorder will produce a more abrupt and stronger absorption than an edge of 60° as found in the flute. The choice of the sharp labium in the recorder might therefore be dictated by the necessity of a stronger production of higher harmonics.

C. Simple source models

1. Jet-drive model

The mouth of the resonator can be described as a two-dimensional slit of width W between a wall and a semi-infinite plane (perpendicular to the wall) representing the labium (Fig. 1). The jet volume flow is assumed to be split into two complementary monopole (line) sources $Q_{\text{in}}=|Q_j|e^{i\omega t}=-Q_{\text{out}}$ placed at a distance ϵ from the edge of the labium at the lower and upper side of the labium, respectively. This jet-drive model has been used by Verge *et al.* (1994a) for small jet thickness h_j compared to the mouth width W. The source is represented by a fluctuating pressure Δp_j deduced from the potential difference across the mouth induced by two monopoles:

$$\Delta p_j \approx \rho_0 \frac{\delta_j}{HW}\frac{dQ_j}{dt}, \tag{9}$$

where δ_j is the effective distance between the two complementary monopole sources Q_{in} and Q_{out}. The length δ_j is calculated by means of a potential flow theory. For the limit $\epsilon \ll W$, we have, following Verge *et al.* (1994a, b),

$$\frac{\delta_j}{W} \approx \frac{4}{\pi}\sqrt{\frac{2\epsilon}{W}}. \tag{10}$$

In the thin jet limit ($W/h \gg 1$), the only length scale in the flow around the edge of the labium is the jet thickness h_j, so that ϵ should be proportional to h_j. Verge *et al.* (1994a, b) actually assume that $\epsilon = h_j$. We make an additional approximation by identifying the jet thickness h_j with the height h of the flue channel. Hence, we neglect the lateral broadening of the jet velocity profile as the jet travels across the mouth.

The power generated by the source is calculated by assuming that the acoustical flow is locally a two-dimensional incompressible flow and that the source is in phase with the acoustical flux through the mouth $Q=(d\xi/dt)WH$ (where WH is the cross-sectional area of the mouth). This corresponds to the condition for which the oscillation amplitude has a maximum as a function of the blowing pressure. This assumption allows us to ignore the effect of the jet response on the phase lag between the sound source and the acoustical flow velocity oscillation. We assume also that for each half oscillation period, the jet direction alternates between inside and outside the resonator. The passage of the jet from one side to the other side of the labium is assumed to be within a very short time compared to the oscillation period. The sources Q_{in} and Q_{out} behave in terms of the time as a square wave of amplitude $Q_j/2$ [Fig. 1(b)]:

$$Q_{\text{in}} = \frac{Q_j}{2} + Q'_{\text{in}}, \quad Q_{\text{out}} = \frac{Q_j}{2} + Q'_{\text{out}}. \tag{11}$$

This model should be reasonable for large displacements of the jet at the labium which are observed in our visualizations for $W/h > 2$.

As explained above, the sources Q_{in} and Q_{out} have an opposite phase ($Q'_{\text{out}} = -Q'_{\text{in}}$) because Q_j is constant ($Q_j = Q_{\text{in}} + Q_{\text{out}}$). The average over an oscillation period T of the power generated by this jet drive is then

$$\begin{aligned}\langle \mathcal{P}_{\text{jet-drive}} \rangle &= \langle \Delta p_j Q \rangle \\ &= \frac{2}{TH}\int_0^{T/2} \rho_0 \frac{4}{\pi}\sqrt{\frac{2h}{W}}|Q_j|\,\delta\!\left(t-\frac{T}{4}\right)\left(\frac{d\xi}{dt}\right)HW\,dt \\ &= \frac{8}{\pi T}\rho_0\sqrt{\frac{2h}{W}}U_0 hHW\left|\frac{d\xi}{dt}\right|,\end{aligned} \tag{12}$$

where U_0 is the jet flow velocity.

By balancing the acoustical energy losses $\langle \mathcal{P}_{\text{losses}} \rangle$ due to vortex-shedding at the labium [Eq. (8)] to the acoustic energy produced by the jet $\langle \mathcal{P}_{\text{jet-drive}} \rangle$ [Eq. (12)], Verge *et al.* (1997a, b) found a relation between the pulsation amplitude $(d\xi/dt)_{\max}$, the Strouhal number $S_r = fW/U_0$, and the ratio W/h of the mouth width W and the jet thickness h:

$$\left(\frac{(d\xi/dt)_{\max}}{U_0}\right)^2 \sim S_r\left(\frac{h}{W}\right)^{3/2}. \tag{13}$$

The agreement of our experimental data with this relationship will be checked in Sec. III (Fig. 11).

While we have assumed that $W/h \gg 1$, we also have assumed that the jet does not break down into discrete vortices. This is only reasonable for the first hydrodynamic mode $S_r = fW/U_0 < 0.3$. As observed by Fletcher (1976), a jet-drive model as proposed here would severely overestimate sound production by higher-order hydrodynamic modes of the jet

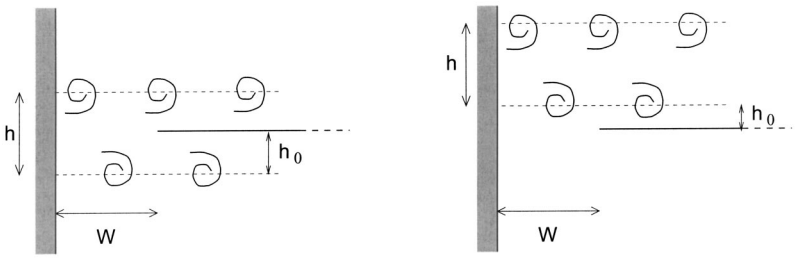


FIG. 2. Idealized representation of the mouth of the instrument in the discrete-vortex model. The two shear layers are separated by a distance h and the parameter h_0 represents the distance between the outer shear layer and the labium.

oscillation. For higher-order modes, a discrete-vortex model should be used. The use of Holger's model (Holger *et al.*, 1977) could be considered here.

2. Discrete-vortex model

When the jet is thick, it will appear from experimental observations that the grazing jet flow is not fully deflected into the resonator. A description in terms of the jet-drive model becomes difficult. If the jet is very thick, the two shear layers bounding the jet behave independently of each other. We follow here the idea of Meissner (2002) to describe both shear layers bounding the jet in terms of discrete vortices. This corresponds to the application of Nelson's model (Nelson *et al.*, 1983; Hirschberg, 1995; Dequand, 2001) to each shear layer. As in the jet-drive model described in the previous section, the mouth of the resonator is described as a two-dimensional slit of width W between a wall and a semi-infinite plane representing the labium (Fig. 2). In Nelson's model (Nelson *et al.*, 1983), the vorticity of the shear layers of the jet is assumed to be concentrated into line vortices traveling along straight lines (in the axis direction of the flue channel) with a constant velocity $U_\Gamma = 0.4U_0$. This value of U_Γ corresponds to experimental observations (Bruggeman *et al.*, 1989). We assume that the circulation of the vortices grows linearly during the first period of their life (before remaining constant).

The two shear layers are separated from each other by a distance h which is assumed to be constant and equal to the jet thickness at the flue exit. In order to avoid the singularity at the labium, the positions of the two shear layers are chosen such that the vortices pass along the labium but do not hit the labium (Fig. 2). We introduce the parameter h_0 corresponding to the distance between the lower shear layer and the labium. The value of h_0 is not critical when varied in the range $W/10 \leqslant h_0 \leqslant W/4$. In the results presented, we used $h_0 = W/5$.

The generation of vortices is periodic and is triggered by the acoustical flux through the mouth. A new vortex is formed at the inner shear layer (on the resonator side) each time the acoustical velocity $d\xi/dt$ changes sign from directed towards the outside to directed towards the inside of the resonator [minimum of the acoustical pressure $p'(L)$ in the resonator]. A new vortex is formed at the outer shear layer half an oscillation period later [maximum of the acoustical pressure $p'(L)$ in the resonator]. The circulation of the jth vortex (with $j=1,2,3,...$) at the inner shear layer ($\Gamma_i^{(j)}$) and at the outer shear layer ($\Gamma_o^{(j)}$) can be written as

$$
\begin{aligned}
&\Gamma_i^{(j)}(t) = \frac{U_0^2}{2}[t-(j-1)T] \quad \text{for } (j-1)T \leqslant t \leqslant jT,\\
&\Gamma_i^{(j)}(t) = \frac{U_0^2}{2}T \quad \text{for } t > jT,\\
&\Gamma_o^{(j)}(t) = -\frac{U_0^2}{2}[t-(2j-1)T/2]\\
&\qquad\qquad \text{for } (2j-1)T/2 \leqslant t \leqslant (2j+1)T/2,\\
&\Gamma_o^{(j)}(t) = -\frac{U_0^2}{2}T \quad \text{for } t > (2j+1)T/2,
\end{aligned}
\tag{14}
$$

where T is the oscillation period.

The acoustical power generated by the vortices is calculated by using Howe's energy corollary (Howe, 1980). The acoustical power averaged over one period of oscillations T is

$$
\langle \mathcal{P}_{\text{discrete-vortex}} \rangle = \frac{-\rho_0}{T}\int_0^T \int_{V_S} (\boldsymbol{\omega}\times\mathbf{v})\cdot\mathbf{u}'\, dV\, dt, \tag{15}
$$

where the volume integration is taken over the source region of volume V_S and where the local acoustical flow velocity $\mathbf{u}'$ represents the unsteady part of the potential flow component of the velocity field $\mathbf{v}$.

The vorticity field $\boldsymbol{\omega} = \boldsymbol{\nabla}\times\mathbf{v}$ takes into account the contribution of each vortex of the shear layers:

$$
\boldsymbol{\omega} = \sum_j [\Gamma_i^{(j)}(t)\,\delta(\mathbf{x}-\mathbf{x}_i^{(j)}(t)) + \Gamma_o^{(j)}(t)\,\delta(\mathbf{x}-\mathbf{x}_o^{(j)}(t))]. \tag{16}
$$

The complex conjugate of the local acoustical flow velocity $\mathbf{u}'$ can be determined by means of potential flow theory (Verge *et al.*, 1994a, b; Dequand, 2001, 2001b):

$$
u'^* = \frac{4}{\pi}\frac{d\xi}{dt}\frac{\zeta}{\zeta^2-1}, \tag{17}
$$

where $\zeta = z/W \pm \sqrt{(z/W)^2-1}$ is either the transformation of the upper half of the ζ-plane into the upper half of the real z-plane (positive sign) or the transformation into points of the lower half of the real z-plane (negative sign) (Fig. 3).

This method takes into account the nonuniformity of the acoustical velocity $\mathbf{u}'$, which is not considered by Meissner (2002) or Hirschberg (1995).

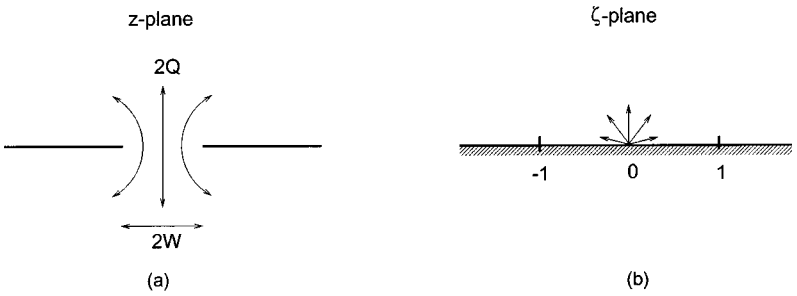


FIG. 3. (a) Idealized geometry deduced from the actual geometry by means of the method of images (z-plane) and (b) transformed plane by means of a conformal mapping (ζ-plane).

3. Global source model

The jet-drive model, valid for thin jets, and the discrete-vortex model, valid for thick jets, can be combined in a global model in which the acoustical power generated by the source is defined as

$$\langle \mathcal{P}_{\text{source}} \rangle = \frac{(W/h)^3 \langle \mathcal{P}_{\text{jet-drive}} \rangle + (W/h)_0^3 \langle \mathcal{P}_{\text{discrete-vortex}} \rangle}{(W/h)^3 + (W/h)_0^3}, \tag{18}$$

where $(W/h)_0$ is a fit parameter. We could choose $(W/h)_0 = 2.5$, which is the value of W/h for which both models (jet-drive and discrete-vortex) give the same source power.

D. Energy balance

Assuming that the particle displacement ξ in the neck is harmonic $\xi(t) = -\hat{\xi} \cos \omega t$, we can write the energy balance of the system from Eqs. (1), (8), and (18). The amplitude $|d\xi/dt|/U_0$ of the oscillations can then be deduced as a function of the Strouhal number $S_r = fW/U_0$.

III. EXPERIMENTAL STUDY

A. Experimental setup

A scheme of the experimental setup is shown in Fig. 4. The resonator is placed in a semi-anechoic room, 60 mm downstream of the outlet (square cross-sectional area of $60 \times 60\ \text{mm}^2$) of a silent wind tunnel. The resonator volume consists out of an aluminum pipe of square cross-sectional area of $H^2 = 60 \times 60\ \text{mm}^2$ and 2 mm wall thickness. Metal plates of 3 mm wall thickness are glued on the pipe walls to reduce potential wall vibrations. The pipe is terminated by a piston in which a piezo-electrical transducer (PCB 116A with Kistler Charge amplifier type 5007) is placed. The piston position can be varied and is determined within 1 mm.

The main pipe, which carries the grazing flow, is terminated 4.9 cm upstream of the neck of the resonator. In the region where the vortices are formed, the side walls of the resonator are made of glass windows.

The mouth consists of three blocks which can be changed to modify the mouth geometry. The instrument is driven by a free jet of thickness h formed by blowing through a slit (flue channel) of height h. This height h was varied between 2 and 60 mm. The slit or flue channel was made by the aperture between a block (lip) of 2.5 cm width and the upstream block of the neck of the cavity. The upstream side of the lip was rounded in order to avoid flow separation within the flue channel, while the downstream side was either sharp [Fig. 5(a)] or chamfered at an angle of 45° [Fig. 5(b)]. Two different angles ($\alpha = -15°$ and 60°) of labium were used. Note that the 15°-labium is directed outside of the cavity while the 60°-labium is directed inside the cavity. These configurations (angle of the labium and direction) correspond to the mouth of a transverse flute and of a recorder, respectively. The distance W between the flue channel exit and the labium was either 24 mm for the 60°-labium or 20 mm for the 15°-labium for a flue exit with sharp edges. With edges, the distance W was increased by 6 mm to 30 and 26 mm, respectively, for the two different labia. In the region surrounding the mouth of the resonator, the side walls of the resonator are made up of glass windows to allow for flow visualization.

B. Acoustical measurements

During the acoustical measurements, the amplitude of the acoustical pressure p'_{exp} at the top of the resonator, the frequency of the acoustical oscillations f, the length L of the resonator, and the jet flow velocity U_0 were determined. As we consider low frequencies, only plane waves propagate in the cavity of the resonator. The mean acoustical velocity amplitude $|d\xi/dt|$ (defined as the ratio between the acoustical flux through the mouth and the mouth opening area $H \times W$) can then be deduced by the formula

$$\frac{d\xi}{dt} = \frac{H}{W} \frac{|p'_{\text{exp}}|}{\rho_0 c_0} \sin\left(\frac{2\pi f L}{c_0}\right) \sin(2\pi f t). \tag{19}$$

1. Experimental results

The measured pulsation amplitudes $|d\xi/dt|/U_0$ as a function of the Strouhal number $S_r = fW/U_0$ are shown in Figs. 6–9 for the four different configurations: the 15°-

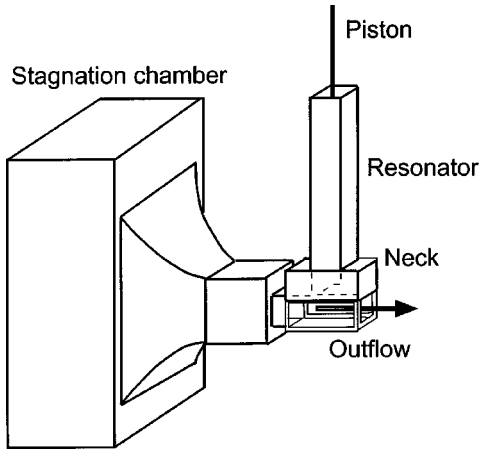


FIG. 4. Experimental setup.

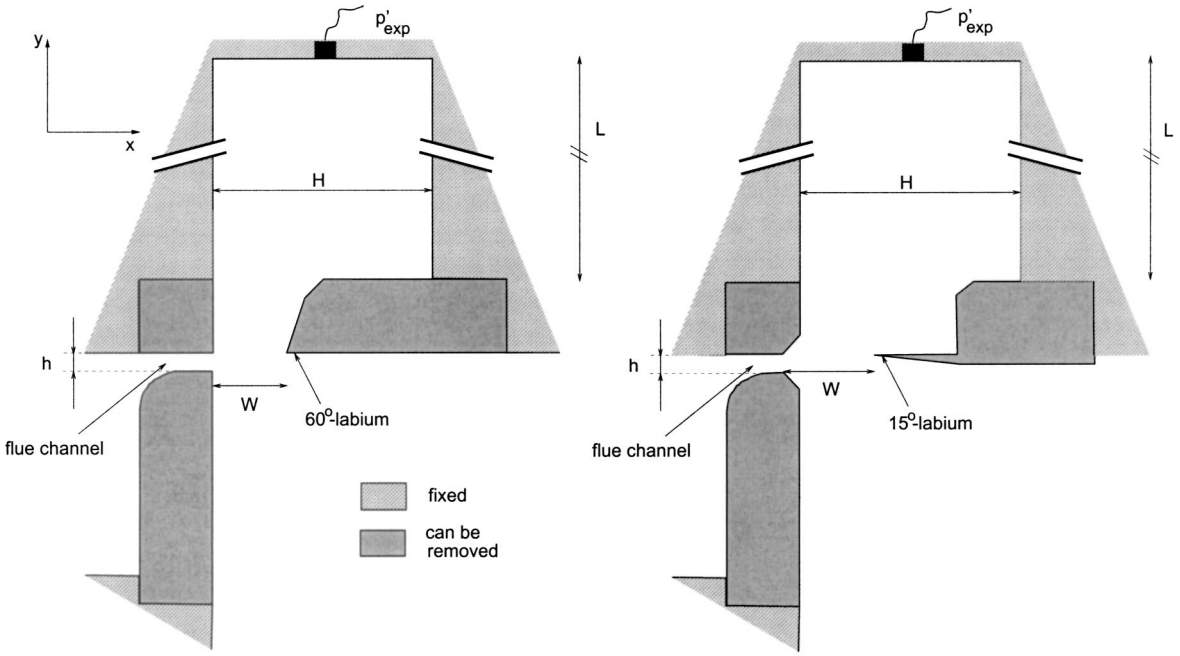
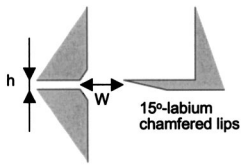


FIG. 5. Mouth geometry. (left) Flue exit with sharp edges and 60°-labium and (right) flue exit with chamfered edges and 15°-labium.

labium and the 60°-labium with either chamfered (Fig. 6 and 8) or sharp lips (Figs. 7 and 9). The results are presented for different heights h of the flue channel. We observe that for thick jets ($h \geqslant 10$ mm), the measured pulsation amplitudes are much less sensitive to the choice of h than for thinner jets.

From the experimental data shown in Figs. 6–9, the maximum of the pulsation amplitudes $(|d\xi/dt|/U_0)_{\max}$ can be deduced and plotted as a function of the ratio W/h of the width W of the mouth opening and the height h of the flue channel. The results are presented in Fig. 10 for the four configurations.



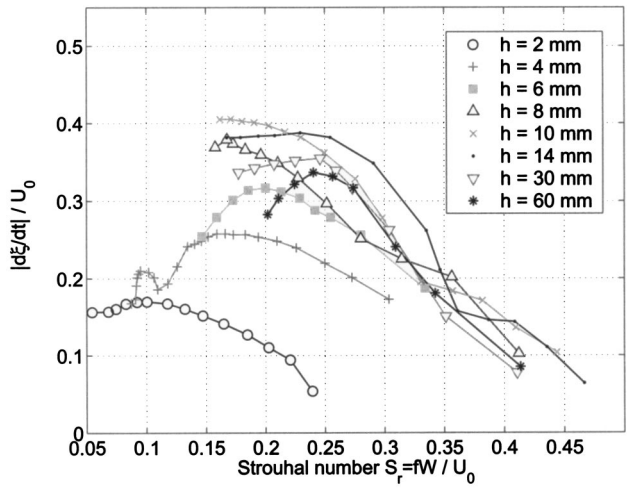
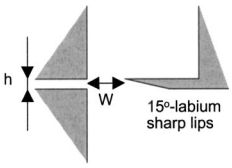


FIG. 6. The 15°-labium with chamfered lips. Pulsation amplitudes $|d\xi/dt|/U_0$ in terms of the Strouhal number $S_r = fW/U_0$ based on the mouth width $W = 26$ mm (length of the resonator $L = 191$ mm).



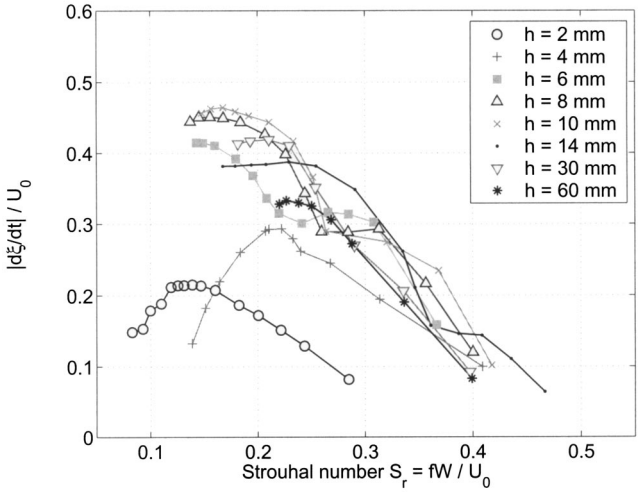
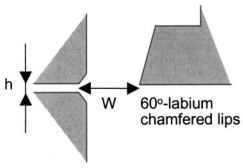


FIG. 7. The 15°-labium with sharp lips. Pulsation amplitudes $|d\xi/dt|/U_0$ in terms of the Strouhal number $S_r = fW/U_0$ based on the mouth width $W = 20$ mm (length of the resonator $L = 191$ mm).



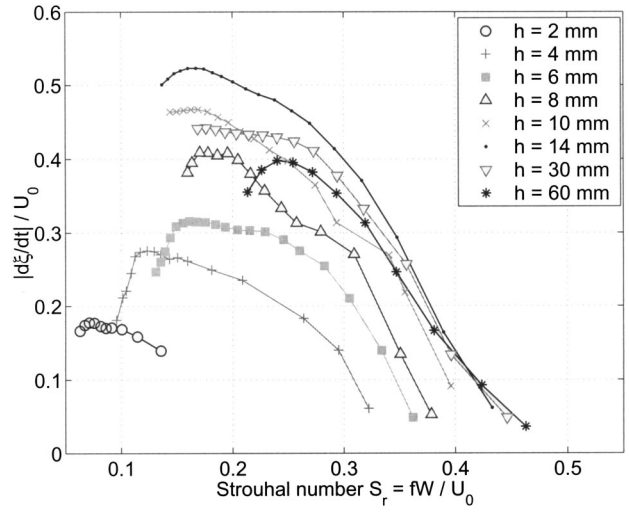
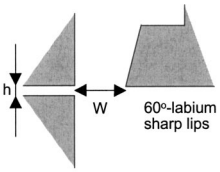


FIG. 8. The 60°-labium with chamfered lips. Pulsation amplitudes $|d\xi/dt|/U_0$ in terms of the Strouhal number $S_r = fW/U_0$ based on the mouth width $W = 30$ mm (length of the resonator $L = 191$ mm).



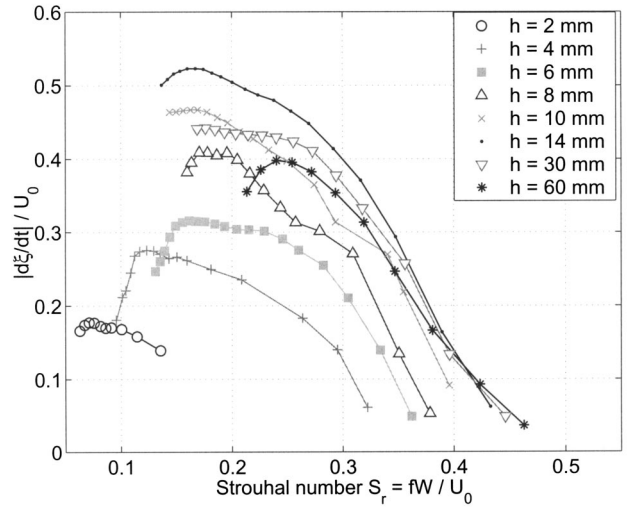


FIG. 9. The 60°-labium with sharp lips. Pulsation amplitudes $|d\xi/dt|/U_0$ in terms of the Strouhal number $S_r = fW/U_0$ based on the mouth width $W = 24$ mm (length of the resonator $L = 191$ mm).

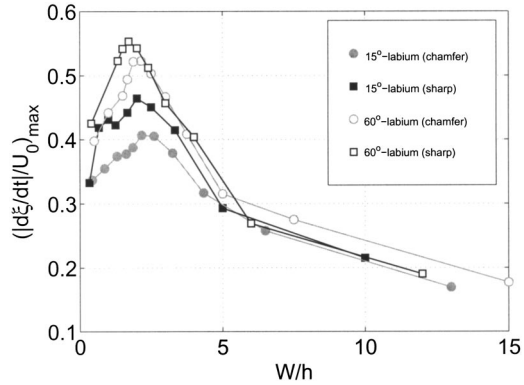


FIG. 10. Maximum of the dimensionless acoustical velocity amplitude $(|d\xi/dt|/U_0)_{\max}$ as a function of the ratio W/h between the width W of the mouth opening and the height h of the flue channel. The data are shown for a flue exit with sharp and chamfered edges and for both the 15° and 60°-labia (length of the resonator $L = 191$ mm).

We observe that the maximum of amplitude $(|d\xi/dt|/U_0)_{\max}$ reached for the 60°-labium is 20% higher than that obtained for the 15°-labium. This difference in amplitude can be explained by the fact that in the case of the 15°-labium, the losses due to vortex shedding at the labium are more important than for the 60°-labium. This difference in amplitude is also observed for real musical instruments as the flute (60°-labium) and the recorder flute (15°-labium). A maximum of the amplitude $(|d\xi/dt|/U_0)_{\max}$ is reached for a ratio W/h around 2. It is interesting to note (Fletcher, 2002) that this corresponds roughly to the maximum of instability for an infinite profile with a Bickley velocity profile (Mattingly and Criminale, 1971) at a hydrodynamical wave number $\pi/(2W)$. Furthermore, when the edges of the flue exit are sharp, the pulsation amplitudes are higher than for a chamfered flue exit. This effect of the geometry of the flue exit has also been observed by Ségoufin *et al.* (2000). In the case of the 15°-labium and a flue exit with sharp edges (Fig. 7), we observe also some instabilities in the measured pulsa-

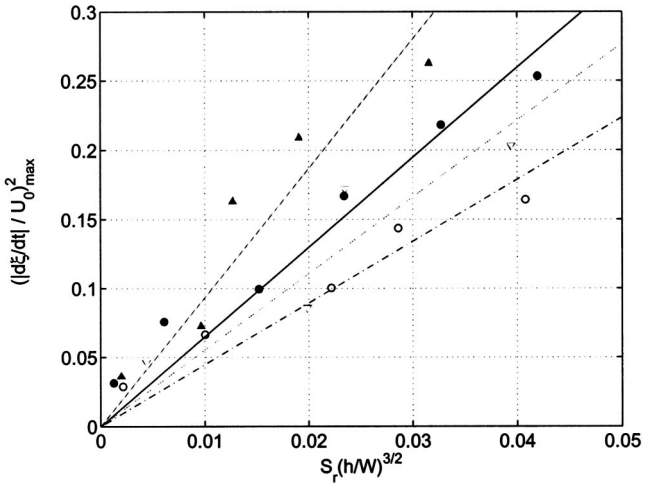


FIG. 11. Jet-drive model for thin jets $(W/h > 2)$. Experimental data (points) are compared with a fit (lines) of the relation found by Verge *et al.* 1997a, b [Eq. (13)] for the four different mouth geometries: 60°-labium and flue exit with chamfered edges (●, ———), 60°-labium and flue exit with sharp edges (▲, ———), 15°-labium and flue exit with chamfered edges (○, ·–·–) and 15°-labium and flue exit with sharp edges (▽, – – –).

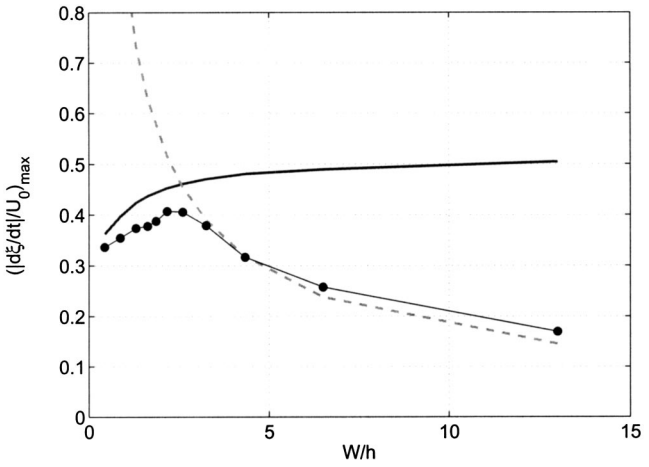


FIG. 12. Maximum of the dimensionless acoustical velocity amplitude $(|d\xi/dt|/U_0)_{\max}$ as a function of the ratio W/h between the width W of the mouth opening and the height h of the flue channel. The data are shown for a flue exit with chamfered edges and for the 15°-labium. Experimental data (●) are compared to the values predicted by means of the proposed analytical models: discrete-vortex model (———) and jet-drive model (– – –). In the discrete-vortex model, the distance between the lower shear layer and the labium is set to $h_0=5$ mm (Fig. 2).

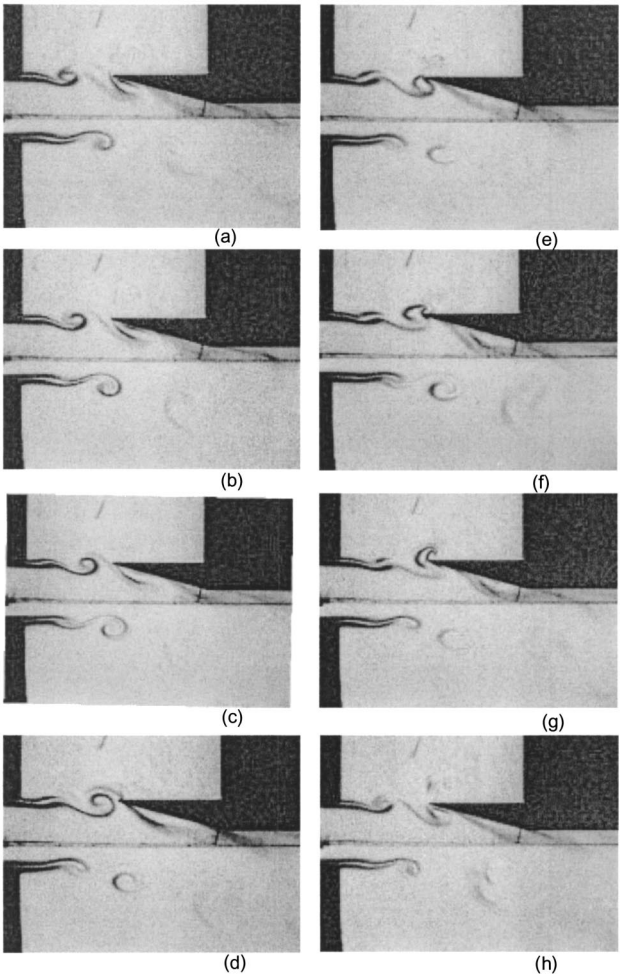


FIG. 13. Flow separation at the 15°-labium of the resonator induced by the passage of vortices near the labium. This behavior was also observed by Verge (1997a, b) for a recorder.

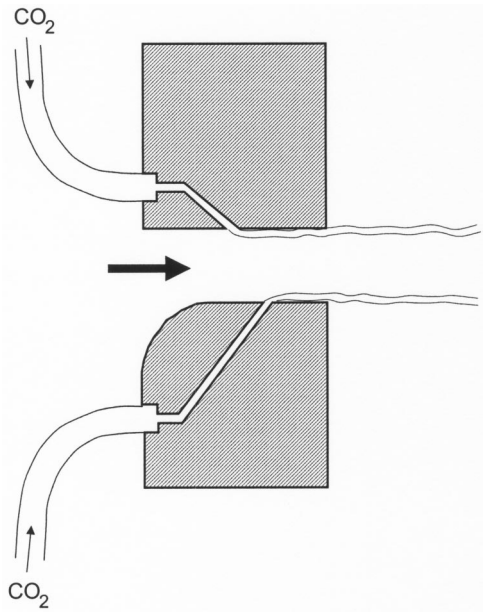


FIG. 14. Scheme of the slit in the blocks of the lips, used for the injection of CO_2 during the flow visualization.

tion amplitudes. This behavior does not occur for the other mouth geometries considered or is much less pronounced.

From Fig. 10, the two types of behavior of the flow in the mouth of the resonator, which were described in Sec. II, are pointed out. For high ratios W/h $(W/h>2)$, the jet formed at the exit of the flue channel is thin and the flow behavior can be described by the jet-drive model presented in Sec. II. For smaller ratios W/h $(W/h<2)$, the jet becomes too thick to remain in one bulk and it breaks down in two ''independent'' shear layers. This is described by the discrete-vortex model (Sec. II).

For thin jets $(W/h>2)$, the relation deduced by Verge *et al.* (1997a, b) from the jet-drive model (Verge *et al.* 1994a) is checked [Eq. (13)] in Fig. 11. The points represent the experimental data and the lines correspond to a least-square fit of the experimental data. We see that the proportionality between $(|d\xi/dt|_{\max}/U_0)^2$ and $S_r(h/W)^{3/2}$ is indeed a fair approximation of the behavior for thin jets. For the 15°-labium in the case of a flue exit with sharp edges (▽), we observe more scatters around this ideal than for the other mouth geometries. As discussed above, the sound production was rather unstable in that particular case.

2. Comparison with the simple model proposed

Figure 12 shows the prediction of the maximum of the pulsation amplitudes $(|d\xi/dt|/U_0)_{\max}$ as a function of the ratio W/h for a flue exit with chamfered edges and the 15°-labium. The analytical data obtained from the simple model described in Sec. II are compared to experimental results.

The model is able to predict the two expected types of behavior as a function of the jet thickness h compared to the mouth width W. This model could be improved by taking into account the flow separation induced at the labium by the passage of vortices (Fig. 13).

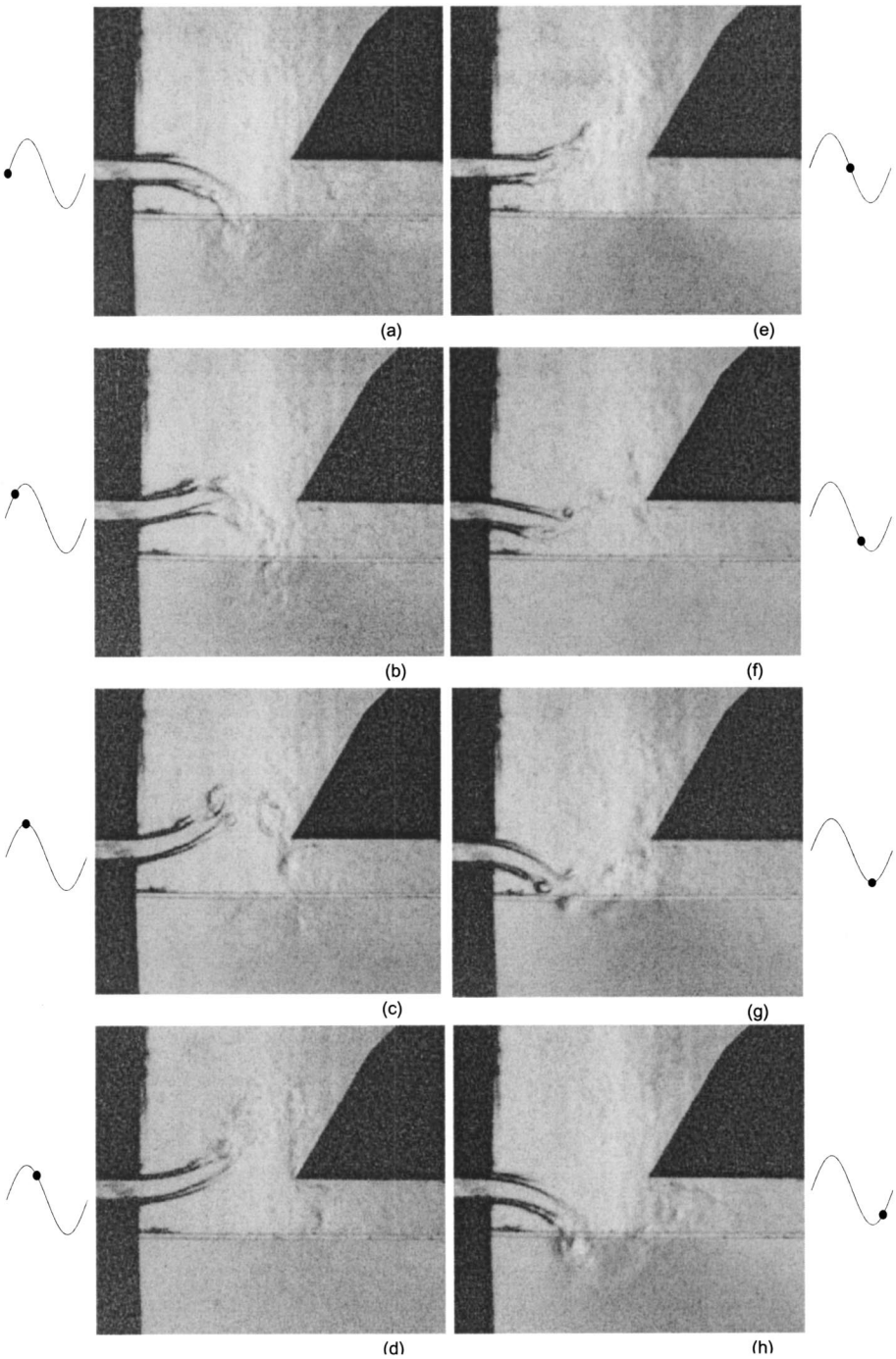


FIG. 15. Vortex shedding in the mouth of the resonator with the 60°-labium, a flue exit with sharp edges, and a channel height $h=4$ mm ($W/h=6$) ($L=552$ mm, $U_0=16.26$ m/s, $S_r=0.19$). The jet swings around the labium and enters entirely into the resonator during part of the oscillation cycle. The time, at which each picture is taken, is shown on the acoustical velocity signal. The pictures (a)–(h) are at $t/T=0.002,0.13,0.25,0.37,0.50,0.63,0.75,0.88$, respectively. The origin $t/T=0$ is the point at which the sign of the acoustical velocity becomes positive.

C. Flow visualization

The periodic flow in the mouth of the resonator is visualized by means of a standard schlieren technique (Merzkirch, 1987).

A lens is placed at each side of the mouth of the resonator. The source of light used is a nanolite light and is placed at the focus length of the first lens so that the beam of light is parallel between the two lenses. The nanolite spark discharge provides a light pulse of about 80 ns duration each time the acoustic pressure at the top of the resonator exceeds a certain value. An additional delay between the triggered signal and the nanolite pulse is introduced in order to visualize the flow at a different moment of the oscillation period T. The injection of a gas with a different refractive index in the mouth of the resonator results in the deviation of the light beam and enables the visualization of the shear layer. At the focus length of the second lens, a small diaphragm is placed in order to be able to adjust the luminosity and the contrast of the pictures. The resulting pictures are taken by means of a camera placed just behind the diaphragm.

Flow visualization has been carried out on a flue exit with sharp edges for the 60°-labium. The gas injected is CO_2. The injection is done by means of a slit in the blocks of the lips (Fig. 14). This slit is made at an angle of 45° in order to reduce disturbances of the shear layers.

The results are shown in Figs. 15–17 for three different

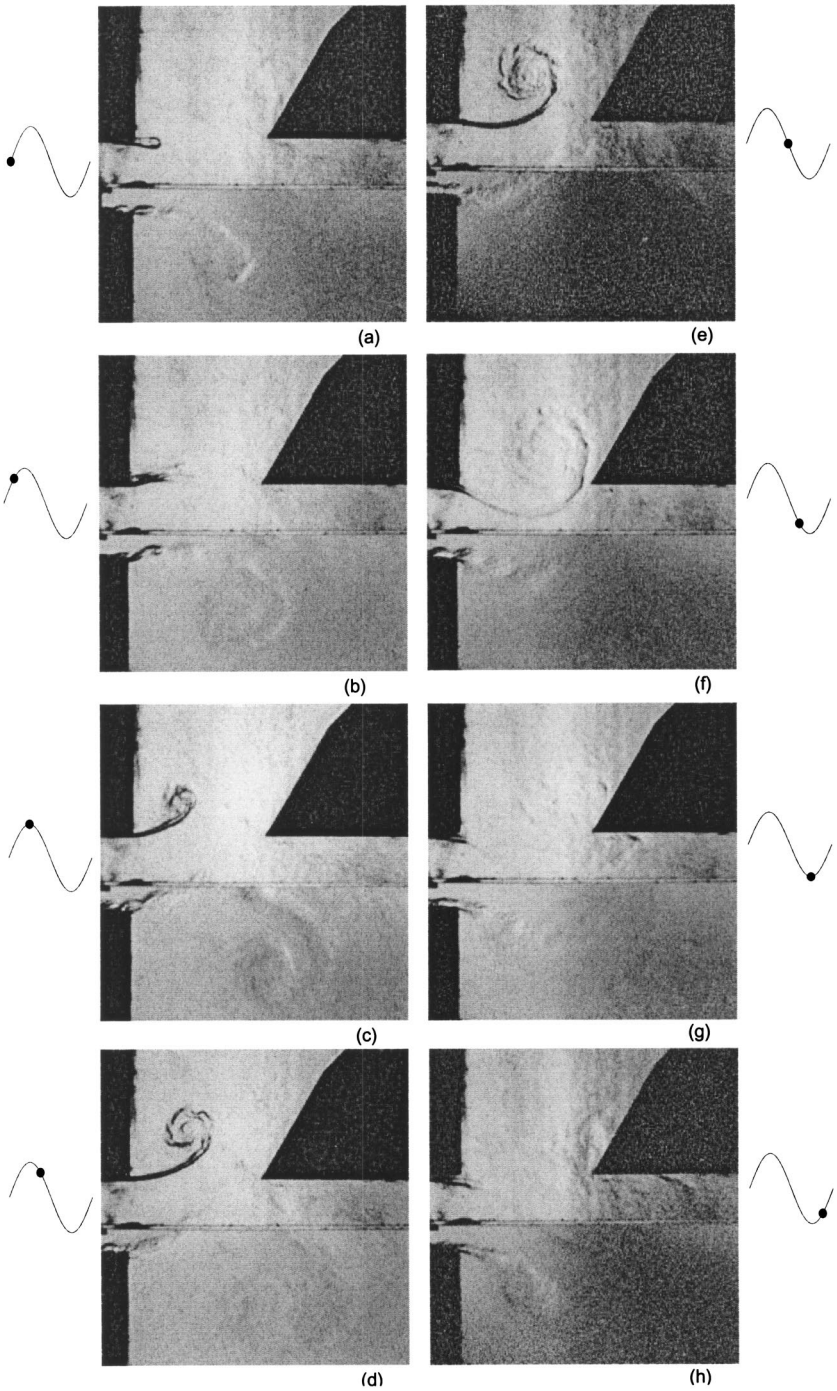


FIG. 16. Vortex shedding in the mouth of the resonator with the 60°-labium, a flue exit with sharp edges, and a channel height $h=14$ mm ($W/h=1.7$) ($L=552$ mm, $U_0=14$ m/s, $S_r=0.22$). We are near the transition $W/h=2$ between the jet-drive behavior and the individual shear layer behavior. The time, at which each picture is taken, is shown on the acoustical velocity signal. The pictures (a)–(h) are at $t/T=0,0.12,0.25,0.37,0.50,0.62,0.76,0.87$, respectively. The origin $t/T=0$ is the point at which the sign of the acoustical velocity becomes positive.

heights h of the channel. For $W/h=6$ (Fig. 15), the jet is thin and oscillates around the labium. When the acoustical pressure at the top of the resonator is minimal, the acoustical velocity starts entering the resonator and the jet is directed into the resonator during half an oscillation period. During the second half of the oscillation period, the acoustical velocity changes sign and the jet is directed out of the resonator.

For $W/h=1.7$ (Fig. 16) and $W/h=0.8$ (Fig. 17), the jet breaks down into two independent shear layers. The lower shear layer does not enter into the resonator. The jet-drive model certainly fails in these cases. When the acoustical velocity is positive (enters the resonator), a first vortex is formed at the upper side of the lips. Half an oscillation period later, a second vortex is formed at the lower side of the lips. Only the first hydrodynamic mode was visualized. For higher Strouhal numbers (smaller main flow velocity), higher hydrodynamic modes can be visualized (Fig. 18).

IV. CONCLUDING REMARKS

The recorder has a relatively poor acoustical response for higher harmonics compared to the transverse flute. We expect that the very sharp labium (15°-labium) of the recorder will generate strong higher harmonics in the source spectrum due to impulsive vortex shedding at the tip of the labium (Verge *et al.*, 1994b; Nolle, 1983). A turbulent jet would in such a case produce a very noisy sound which is

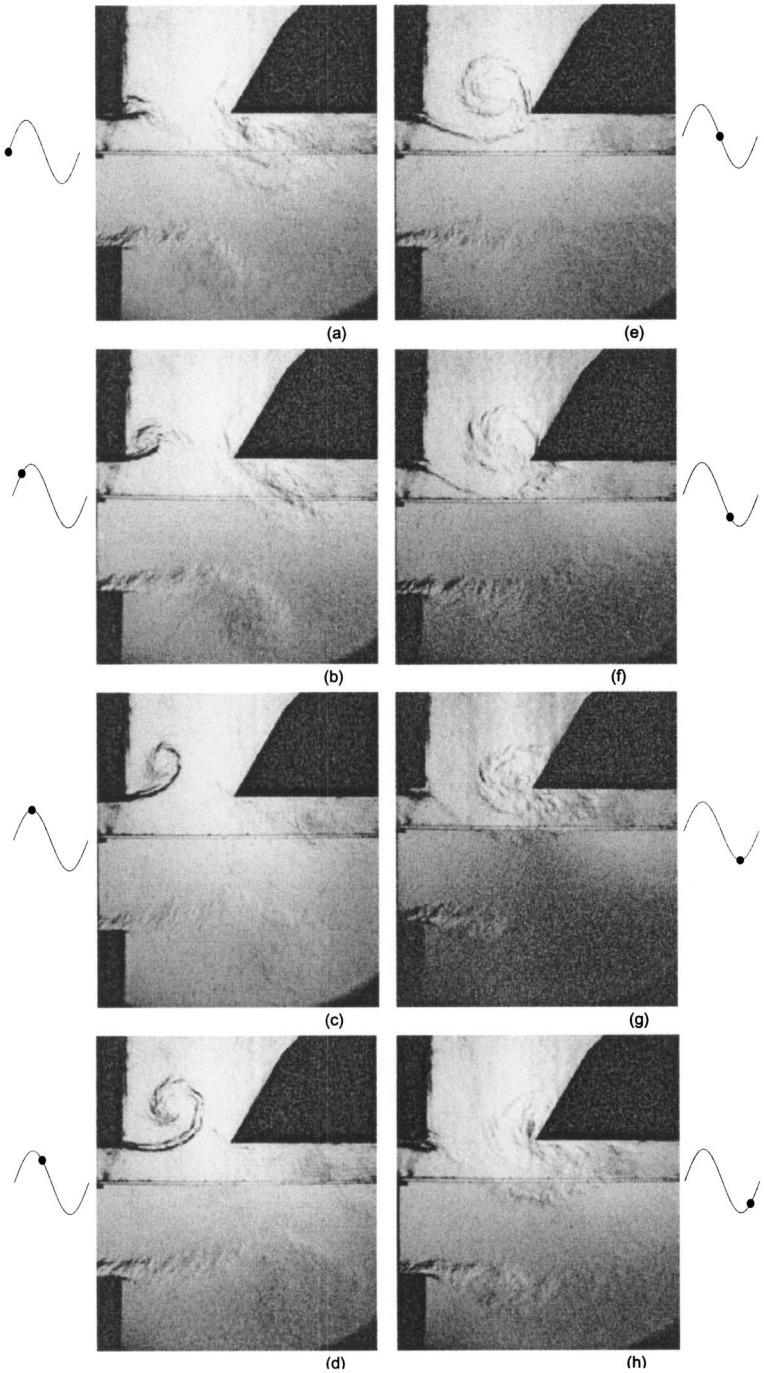


FIG. 17. Vortex shedding in the mouth of the resonator with the 60°-labium, a flue exit with sharp edges, and a channel height $h=30$ mm ($W/h=0.8$) ($L=552$ mm, $U_0=14.5$ m/s, $S_r=0.22$). The outer shear layer remains far away from the opening of the resonator. The time, at which each picture is taken, is shown on the acoustical velocity signal. The pictures (a)–(h) are at $t/T=0,0.12,0.25,0.35,0.45,0.6,0.75,0.85$, respectively. The origin $t/T=0$ is the point at which the sign of the acoustical velocity becomes positive.

avoided by keeping the blowing pressure of the instrument low. This is at the expense of the power of the instrument. In a transverse flute (Boehm flute), the excellent harmonicity of higher modes results in a rich sound even for a thick labium with an angle of 60°. This has the advantage of a reduction of turbulent noise at a given blowing pressure but also appears to provide a more stable jet oscillation for thick jets. As the flute operates with a variable jet thickness (lips of the musician), this could be an additional argument to use a labium with a large angle (60°).

Two simplified source models for flue instruments have been proposed. The models explain the variation of the oscillation amplitudes as a function of the ratio W/h of the mouth width W to the jet thickness h for low Strouhal numbers. The jet-drive model (Coltman, 1976) is valid for thin jets ($W/h>2$) which are typically present in musical instruments. A discrete-vortex model, inspired by the works of Holger *et al.* (1977) and Meissner (2002), has been proposed to explain the sound production of the instrument when the jet-drive model fails ($W/h<2$). These two models can be combined in a global model which could be used in a sound synthesis model if the jet-drive model is complemented with a model predicting the effect of the sound source relative to the acoustical oscillation. In first approximation, a simple delay line could be considered. Alternatively, the approach of Fletcher and Rossing (1998) could be used. It is furthermore clear that our model will not predict accurately the spectral envelope of the sound produced by the instrument. Finally,

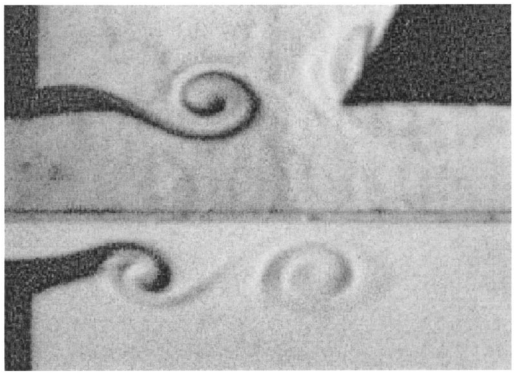


FIG. 18. Visualization of the second hydrodynamic mode in the mouth of the resonator with the 60°-labium, a flue exit with sharp edges, and a channel height $h=14$ mm ($W/h=1.7$) (length of the resonator: $L=552$ mm).

we would like to mention that we are now studying the effect of the acoustics of the instrument on the sound source. Preliminary results show no major effects of the acoustics on the flow.

Brown, C. E., and Michael, W. H. (**1954**). ''Effect of Loading Edge Separation on the Lift of a Delta Wing,'' J. Aeronaut. Sci. **21**, 690–695.

Bruggeman, J. C., Hirschberg, A., van Dongen, M. E. H., Wijnands, A. P. J., and Gorter, J. (**1989**). ''Flow Induced Pulsations in Gas Transport Systems: Analysis of the Influence of Closed Side Branches,'' J. Fluids Eng. **111**, 484–491.

Campbell, M., and Greated G. (**1987**). *The Musician's Guide to Acoustics* (Schirmer, New York).

Castellengo, M. (**1976**). ''Contribution à l'étude expérimentale des tuyaux à bouche,'' Ph.D. thesis, Université de Paris VI, Paris.

Coltman, J. W. (**1966**). ''Resonance and sounding frequencies of the flute,'' J. Acoust. Soc. Am. **40**, 99–107.

Coltman, J. W. (**1968**). ''Sounding mechanism of the flute and organ pipe,'' J. Acoust. Soc. Am. **44**, 983–992.

Coltman, J. W. (**1969**). ''Sound radiation from the mouth of an organ pipe,'' J. Acoust. Soc. Am. **46**, 477.

Coltman, J. W. (**1976**). ''Jet drive mechanism in edge tones and organ pipes,'' J. Acoust. Soc. Am. **60**, 725–733.

Coltman, J. W. (**1980**).

Crighton, D. G. (**1992**). ''The Edge Tone Feedback Cycle. Linear Theory for the Operating Stages,'' J. Fluid Mech. **234**, 361–391.

Dequand, S. (**2001**). ''Duct Aeroacoustics: from Technological Applications to the Flute,'' Ph.D. thesis, Technische Universiteit Eindhoven, Netherlands and Université du Maine, France.

Dequand, S., Luo, X., Willems, J. F. H., and Hirschberg, A. (**2001**). ''Self-Sustained Oscillations in a Helmholtz-like Resonator. Part 1: Acoustical Measurements and Analytical Models,'' in *7th AIAA/CEAS Aeroacoustics Conference* (Maastricht, Nl), submitted to the AIAA Journal.

Elder, S. A. (**1973**). ''On the mechanism of sound production in organ pipes,'' J. Acoust. Soc. Am. **54**, 1554–1564.

Elder, S. A. (**1992**). ''The Mechanism of Sound Production in Organ Pipes and Cavity Resonators,'' J. Acoust. Soc. Jpn. (E) **13**, 11–23.

Fabre, B., and Hirschberg, A. (**2000**). ''Physical Modelling of Flue Instruments: a Review of Lumped Models,'' Acust. Acta Acust. **86**, 599–610.

Fabre, B., Hirschberg, A., and Wijnands, A. P. J. (**1996**). ''Vortex Shedding in Steady Oscillations of a flue Organ Pipe,'' Acust. Acta Acust. **82**, 863–877.

Fletcher, N. (**1976**). ''Jet drive mechanism in organ pipes,'' J. Acoust. Soc. Am. **60**, 481–483.

Fletcher, N. (**2002**). Private communication.

Fletcher, N., and Rossing, T. (**1998**). *The Physics of Musical Instruments* (Springer-Verlag, New York).

Hirschberg, A. (**1995**). ''Aeroacoustics of Wind Instruments'' in *Mechanics of Musical Instruments*, International Centre for Mechanical Sciences, Courses and Lectures No. 35, edited by Hirschberg A., Kergomard J., and Weinrich (Springer-Verlag, Wien), pp. 291–369.

Holger, D. K., Wilson, T. A., and Beavers, G. S. (**1977**). ''Fluid mechanics of the edge-tone,'' J. Acoust. Soc. Am. **62**, 1116–1128.

Howe, M. S. (**1975**). ''Contributions to the Theory of Aerodynamic Sound, with Application to Excess Jet Noise and the Theory of the Flute,'' J. Fluid Mech. **71**(4), 625–673.

Howe, M. S. (**1980**). ''The Dissipation of Sound at an Edge,'' J. Sound Vib. **70**, 407–411.

Howe, M. S. (**1998**). *Acoustics of Fluid-Structure Interactions* (Cambridge U.P., Cambridge).

Ingard, K. U., and Ising, H. (**1967**). ''Acoustic nonlinearity of an orifice,'' J. Acoust. Soc. Am. **42**, 6–17.

Mattingly, G., and Criminale, W. (**1971**). ''Disturbance characteristics in a plane jet,'' Phys. Fluids **14**, 2258–2264.

Meissner, M. (**2002**). ''Aerodynamically Excited Acoustic Oscillations in Cavity Resonator Exposed to an Air Jet,'' Acust. Acta Acust. **88**, 170–180.

Merzkirch, W. (**1987**). *Flow Visualization*, 2nd ed. (Academic, London).

Nederveen, C. J. (**1998**). *Acoustical Aspects of Woodwind Instruments*, 2nd ed. (Northern Illinois U.P., De Kalb).

Nelson, P. A., Halliwell, N. A., and Doak, P. E. (**1983**). ''Fluid Dynamics of a Flow Excited Resonance. Part 2: Flow Acoustic Interaction,'' J. Sound Vib. **91**, 375–402.

Nolle, A. W. (**1983**). ''Flue organ pipes: Adjustments affecting steady waveform,'' J. Acoust. Soc. Am. **73**, 1821–1832.

Peters, M. C. A. M. (**1993**). ''Aeroacoustic Sources in Internal Flows,'' Ph.D. thesis, Technische Universiteit Eindhoven.

Powell, A. (**1961**). ''On the edge-tone,'' J. Acoust. Soc. Am. **33**, 395–409.

Pullin, D. I. (**1978**). ''The large scale structure of unsteady self-similar rolled-up vortex sheets,'' J. Fluid Mech. **88**, 401–430.

Ségoufin, C., Fabre, B., Verge, M. P., Hirschberg, A., and Wijnands, A. P. J. (**2000**). ''Experimental Study of the Influence of the Mouth Geometry on Sound Production in a Recorder-like Instrument: Windway Length and Chamfers,'' Acust. Acta Acust. **86**, 649–661.

Ségoufin, C., Fabre, B., and Thiria, B. (**2001**). ''Experimental Confrontation of Linear and Non-Linear Jet Theories Production in Recorder Type Instrument,'' in *Proceedings of the International Symposium on Musical Acoustics ISMA*, Vol. 1, Perugia, Italy.

Verge, M. P., Caussé, R., Fabre, B., Hirschberg, A., Wijnands, A. P. J., and van Steenbergen, A. (**1994**). ''Jet Oscillations and Jet Drive in Recorder-like Instruments,'' Acta Acust. united with Acustica **2**, 403–419.

Verge, M. P., Fabre, B., and Hirschberg, A. (**1994b**). ''Jet formation and jet velocity fluctuations in a flue organ pipe,'' J. Acoust. Soc. Am. **95**, 1119–1132.

Verge, M. P., Fabre, B., Hirschberg, A., and Wijnands, A. P. J. (**1997a**). ''Sound production in a recorder-like instrument. Part 1: Dimensionless amplitude of the internal acoustic field,'' J. Acoust. Soc. Am. **101**, 2914–2924.

Verge, M. P., Fabre, B., Hirschberg, A., and Wijnands, A. P. J. (**1997b**). ''Sound production in a recorder-like instrument. Part 2: A simulation model,'' J. Acoust. Soc. Am. **101**, 2925–2939.

Wolfe, J., Smith, J., and Fletcher, N. H. (**2001**). ''Acoustic Impedance Spectra of Classical and Modern Flutes,'' J. Sound Vib. **243**, 127–144.

Musical quality assessment of clarinet reeds using optical holography

Fabrice Pinard,[a] Benoit Laine,[a] and Holger Vach[b]
Laboratoire de Physique des Interfaces et des Couches Minces, CNRS–Ecole Polytechnique, F-91128 Palaiseau Cedex, France

(Received 11 December 2001; revised 13 December 2002; accepted 16 December 2002)

Vibrational modes of 24 clarinet reeds have been observed in both dry and wet conditions using holographic interferometry. Results have been compared with the ''musical quality'' of the reeds as judged by two professional clarinet players. An excellent correspondence has been demonstrated between specific vibrational behavior and musical quality. The results suggest that the presence and symmetry of a strong first torsional mode are indicative of good or very good musical quality. A second, but less stringent quality criterion is the proximity of frequencies corresponding to the second torsional and the second flexural mode. This proximity leads to the creation of mixed vibrational modes for the very best of the investigated clarinet reeds. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1543586]

PACS numbers: 43.75.Pq, 43.75.Ef, 43.75.Yy [NHF]

I. INTRODUCTION

The clarinet is a single-reed woodwind instrument. The reed is a small slice of natural cane or synthetic material put into vibration when the musician is blowing air into the instrument. The important role of the lowest frequency reed resonance in sound production by single-reed instruments has been well established.[1] The acoustical significance of the reed's higher vibrational modes has been partly identified for a clarinet reed.[2,3] Holography has been used to locate the corresponding frequencies and to image their vibrational patterns. Some vibrational modes were reported that are similar to the one-dimensional modes of a cantilevered beam. Other modes showed two-dimensional vibrations indicating some twisting motion.[3]

To the best of our knowledge, however, no in-depth study has yet been reported relating the details of vibrational reed behavior to musical quality; i.e., there is no explicit information of how the vibrational mode patterns differ for reeds of different quality. For a professional musician, it is a rather simple and straightforward task to intuitively judge something as complex as ''musical quality.'' For a scientist, however, the same task becomes infinitely more complicated since it is far from trivial to relate musical quality to simple observable quantities. Obviously, it is not only the tip thickness, the reed flexibility, or the cane density that will determine the sound quality. The homogeneity of the reed material should certainly play a crucial role. However, how can nonlinear effects and details of the player's embouchure be considered in sound production and in musical quality measurements? In addition, it is known that a reed only obtains its optimal sound characteristics after a certain amount of use. In the absence of answers to most of those questions, reed manufacturers only test the global stiffness of their reeds in order to classify them before putting them on the market.

As a result, the selection of a clarinet reed becomes a daily problem for any clarinet player around the world. The normal use is to buy several boxes containing 10 reeds each and to hope that it is possible to find a couple of reeds that will give a satisfactorily good sound. Thereafter, of course, new trouble may rapidly occur if these ''good reeds'' experience changes in temperature or wetness, or simply as a result of aging due to being intensively played. The goal of the present study is, therefore, to find out if there are any elementary and reproducible criteria to judge the musical quality of clarinet reeds based on simple and objective measurements. In the following, an in-depth study in very close collaboration with members of professional orchestras will be presented and it will be shown that it is, indeed, possible to predict the normally intuitively determined musical quality from details of the vibrational behavior of the clarinet reeds as it can be visualized using, for example, holographic interferometry.

II. EXPERIMENT

The initial working hypothesis for this study was simply to assume that the ''musical quality'' of clarinet reeds is directly related to their vibrational behavior. In this manner, the professional clarinetists first classified a sample set of 24 reeds in four categories of musical quality ranging from ''very poor'' to ''very good.'' Then, the four or five lowest natural frequencies and vibrational mode patterns for each of those classified clarinet reeds were experimentally determined to examine whether or not it is possible to establish a simple relationship between details of vibrational properties and musical quality.

[a]Present address: Laboratoire Analyse des Matériaux et Identification Ecole Nationale des Ponts et Chaussées, F-77455 Marne la vallée, France. Electronic mail: fabrice.pinard@polytechnique.org
[c]Electronic mail: vach@leonardo.polytechnique.fr

 0001-4966/2003/113(3)/1736/7/$19.00

A. Musical quality evaluation

To get a reliable classification of the clarinet reeds from a musical point of view, the expertise of two professional clarinet players from two of the finest standards in the world has been requested: the first one, Jerome Veragh, is super soloist with the *French National Opera*, and the second one, Arnaud Leroy, is soloist with the *Orchestre de Paris*. Considering that the first expert is used to playing V12 clarinet reeds from Vandoren with a stiffness grading of $3\frac{1}{2}$, and that this is a widespread clarinet reed, we decided to pick a 24-sample set among four different boxes of those Vandoren reeds bought in a regular music-accessories store. Both musicians determined the musical quality of each reed in the set; i.e., they evaluated the resulting sound quality of their instruments as well as the ability and ease with which they could make the reeds vibrate along the whole spectrum of the clarinet. Both musicians tested the reeds using their own instruments, which were in fact identical: the RC model from Buffet Crampon. Accidentally, their mouthpieces were also the same: Vandoren model B40 with standard ligatures. The reeds were not scraped or modified in any way by either one of the musicians.

One very satisfying result was the fact that both musicians established quite the same classification for the provided reeds. This finding is very satisfactory since the second clarinetist (Arnaud Leroy) has quite different playing ''habits'' and normally uses clarinet reeds of the regular model from Vandoren (stiffness number $3\frac{1}{2}$, which is considered less stiff than the equivalent V12-stiffness number). To some extent, he was to carry out a counterevaluation of the previous classification by his colleague. In fact, he proved to be able to evaluate perfectly both ability and ease in reed vibration, even though the reeds were often too stiff for him to play them for a very long time. From the initial 24-reed sample set, our experts agreed to classify four reeds as ''very poor,'' nine as ''poor,'' seven as ''good,'' and four as ''very good.''

B. Holographic setup

The total vibrating area of a clarinet reed fixed to the mouthpiece by its ligature is about 3 to 4 square centimeters, and the weight of the moving part in sound production is just a fraction of a gram. Due to this very low weight, the use of accelerometers becomes impossible and a natural choice for the experimental study is to consider holographic interferometry that combines the advantages of providing a highly refined representation of vibrational mode patterns and of being a noncontact technique.[4]

The employed optical setup is shown in Fig. 1. Isolation from external vibrations was assured by placing the ''homemade'' holographic table on three properly inflated motorcycle inner tubes. The coherent light source was a 5-mW He–Ne laser. A standard holographic Mach–Zehnder configuration with simple, standard optical components has been used to record the holograms. The reference beam was purified and expanded by a spatial filter consisting of a $\times 40$ microscope lens and a 10-μm pinhole. The object beam was expanded by a diffuse glass and its length was chosen to equal the optical path length of the reference beam at the holographic plate.

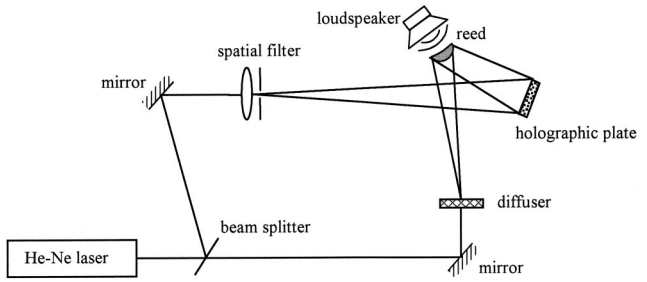


FIG. 1. Sketch of the experimental setup that was used for the holographic interferometry measurements.

To record vibrational mode patterns of the reeds, the time average method was used.[4] For a sinusoidal movement with a maximum amplitude A of the object, the intensity distribution over the surface of the reconstructed object in the hologram, when the illumination time is much larger than the vibrational period, is given by:[4] $I(x,y) \propto J_0^2[2\pi AF(x,y) \times(\cos\varphi_1+\cos\varphi_2)/\lambda]$ where λ is the wavelength of the laser beam, J_0 is the zeroth-order Bessel function of the first kind, φ_1 is the angle between the direction of illumination and the direction of the displacement, φ_2 is the angle between the direction of displacement and the direction of observation, and $F(x,y)$ describes the vibrational mode. Superposed on the hologram of the reed appears, thus, a set of dark and bright interference fringes following regions with equal vibrational amplitude. Nodal lines separate regions that are vibrating in phase opposition. The number of fringes is proportional to the amplitude of the motion. For most holograms, AGFA 8E75HD holographic plates were used and the optimal exposure time for this setup was determined by trial and error to be around 40 to 45 seconds. This means that the holograms were time averaged over roughly 100 000 cycles. Even for those relatively long exposure times, the simple vibrational isolation described above was totally sufficient. The vibrational patterns depicted in the following were obtained by photographing the virtual images of the recorded holograms.

C. Vibrational excitation procedure

As previously discussed, in the same way that the recording technique had to be a noncontact one, any classical method based on a direct mechanical excitation had to be avoided, as it would have introduced a highly dominating mechanical impedance. Therefore, a system providing a pure acoustical excitation was employed. To this end, the output of a low-frequency generator was amplified by a 50-W power amplifier that was connected to a high-fidelity loudspeaker. The only nonstandard arrangement was the addition of a 105-mm-diameter polypropylene tube glued on the wooden part of the speaker box just in front of the 100-mm-diameter cone to achieve two main goals: first, to concentrate the acoustical pressure on the reed during the exposure as the open end of the tube was only 5 mm away from the reed, and second, to drastically reduce perturbations on the holography

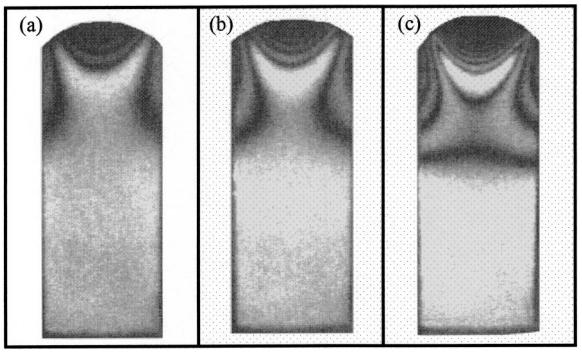


FIG. 2. Vibration patterns obtained with the same clarinet reed excited at the same vibrational resonance frequency, but at three different levels of acoustical excitation: (a) 10 W; (b) 20 W; and (c) 40 W approximate speaker output power.

table due to sound reflection off the walls of the experimental room. Instead of designing a special clamping device, a simple clarinet mouthpiece that offered an easy, safe, and reproducible way to hold the reeds in place was used for the recording of the time-averaged holograms. At the same time, this choice permitted the two professional clarinet players to play the reeds at any time again for further musical testing.

To find the natural reed frequencies, the reflection of a weak He–Ne laser beam from the reed surface on a white screen was observed while slowly scanning the spectrum of excitation frequencies. As soon as a resonance frequency was approached, this image became significantly perturbed. Due to both eye sensitivity and the maximum power of the amplifier, we concentrated on frequencies in the range from 400 to 7000 Hz in the present study. Once a resonance frequency was detected, the corresponding vibrational mode pattern was recorded using the holography setup described above. The excitation level was controlled by simply varying the output power of the amplifier. In Fig. 2, the same reed at the same excitation frequency, but at three different levels of acoustical pressure is depicted: the maximum vibrational amplitude; i.e., the number of interference fringes, increases about linearly with acoustical excitation power.

III. RESULTS AND DISCUSSION

A. Dry reed conditions

A first series of experiments was carried out under dry reed conditions. Even if reeds are obviously wet when being played by a clarinetist, our goal was to investigate whether it was possible to predict the sound quality—as judged by a musician under realistic wet playing conditions—directly from dry reeds. In case the musical quality was more directly related to the vibrational mode patterns than to the absolute value of natural frequencies, it may not be required to test reeds in wet conditions that, of course, are much more complex to control during experiments.

All investigated reeds exhibited a first vibrational mode in the range of frequencies from 2100 to 2400 Hz (see Table I) with a shape similar to the first flexural mode of a cantilevered beam. There was no significant difference between the different reeds either in frequency, or in vibration pattern

TABLE I. First four resonance reed frequencies determined for the 24-reed sample set representing four different musical quality categories as classified by two professional musicians. The cited frequencies were all measured under dry reed conditions. $\Delta\nu_2$ is the measured frequency difference between the second flexural and the second torsional mode. The resonance quality is expressed by the use of different fonts: *hardly detectable*—good—**very strong**.

	1st flexural (Hz)	1st torsional (Hz)	2nd flexural (Hz)	2nd torsional (Hz)	Mixed mode (Hz)	$\Delta\nu_2$ (Hz)
Very poor	**2140**		**5800**	6260		460
Very poor	2220	*3750*	5450	**5830**		380
Very poor	**2200**		**5830**	6370		540
Very poor	**2320**		**5840**	6280		440
Poor	**2285**	*3780*	**5720**	**6250**		530
Poor	**2205**		**5810**	**6270**		460
Poor	**2310**	*3610*	5850	**6200**		350
Poor	**2270**		**5830**	6100		270
Poor	**2200**	*3550*	**5860**	**6350**		490
Poor	**2260**	*3740*	**5880**	*6120*		240
Poor	2380		6390	6710		320
Poor	**2440**		**6315**	6680		365
Poor	**2330**	*3760*	**6320**	**6790**		470
Good	**2320**	**3770**	**5820**	6210		390
Good	**2325**	**3760**	**6325**	**6560**		235
Good	**2240**	3750	**5860**	6100		240
Good	**2330**	**3560**	5930	**6270**		340
Good	**2140**	3750	**5850**		5950	100
Good	2200	3600	5790			
Good	2240	3515	**5890**		5955	65
Very good	**2250**	**3760**	**5980**	**6260**		280
Very good	**2260**	**3600**	**5880**	**6130**		250
Very good	**2320**	**3540**	**5910**	**6200**		290
Very good	**2320**	**3710**			**6220**	<65

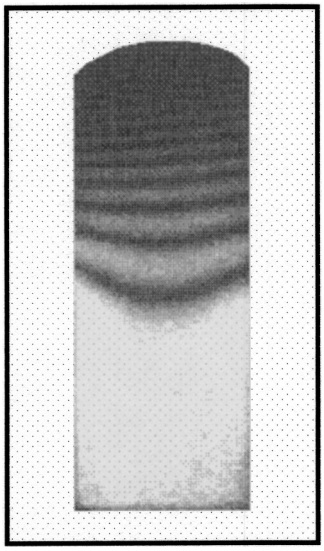


FIG. 3. Typical vibration pattern corresponding to the first flexural mode obtained under dry conditions representative for all reeds independent of their musical quality.

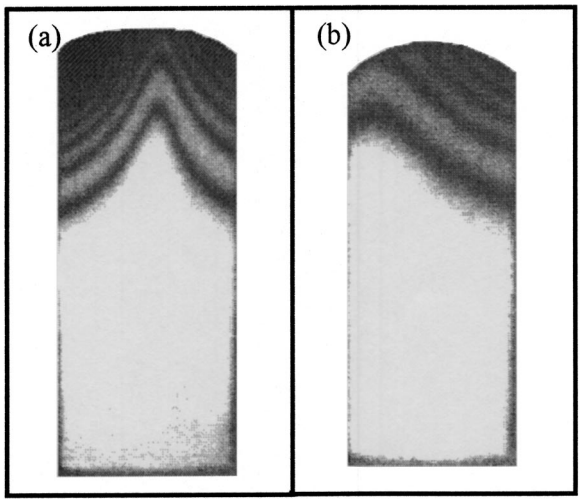


FIG. 4. Typical vibration patterns corresponding to the first torsional mode obtained under dry conditions: (a) slightly non-symmetric pattern for a very good reed and (b) significantly nonsymmetric pattern for a good reed. While the degree of symmetry of the first torsional mode can be used to predict whether a reed has a good or very good quality, the absence of this mode is a clear indication for a reed of poor or even very poor sound quality.

or resonance quality. The picture in Fig. 3 can be considered as a typical one for all reeds independent of their musical quality. The curvature of the interference lines close to the base of the reed is a direct consequence of the boundary condition induced by the form of the contact zone in the vibrating area between the plane face of the reed and the clarinet mouthpiece.

When slowly increasing the excitation frequency, only two types of reeds, the ''good'' and the ''very good'' ones, clearly exhibited another marked resonance between 2500 and 5000 Hz (see the frequencies with bold-type characters in Table I). The observed pattern is very similar to that of the first torsional mode of a cantilevered plate. This mode seems to be highly sensitive to the geometrical and mechanical reed symmetry: even for very good reeds, the displacement amplitude of the left corner can be about twice the one of the right corner according to the number of interference lines that are readily counted in the example of Fig. 4(a). Good reeds might even demonstrate vibrational amplitudes that are about four times higher on one side than on the other [see Fig. 4(b)]. We would like to insist on the fact that neither ''very poor'' nor ''poor'' reeds showed a significant resonance for the first torsional mode; i.e., our present excitation scheme was not sufficiently strong to excite this mode markedly for poor quality reeds, while good and very good reeds always demonstrated a strong resonance for the corresponding frequencies even with relative low excitation power.

Going to yet higher excitation frequencies, modes that correspond to the second torsional and second flexural modes of a simply cantilevered plate could be clearly identified. For both poor and very poor reeds, the corresponding frequencies were located in the range from 5700 to 6800 Hz (see Table I). Figure 5(a) shows a typical vibration pattern of a second torsional mode, while Fig. 5(b) shows the picture of a second flexural mode. For most of the good and very good reeds, the same two vibrational patterns were found in this frequency range. However, two of the seven good reeds exhibited a slightly different vibrational behavior. The lower vibrational resonance was still the image of a pure second flexural mode as the one shown in Fig. 5(b). The other resonance, however, was the image of a mixed mode pattern resulting from a combination of torsional and flexural movement [see Fig. 6(a)]. Among the set of very good reeds, we like to point out our result for the reed that was unanimously classified by both experts as the very best one: this reed only exhibited one single resonance in the range from 5000 to 7000 Hz; i.e., it was impossible to find two distinctively different vibrational modes corresponding to pure second torsional and pure second flexural modes. Instead, only one strong resonance was observed corresponding to a vibrational pattern that again resulted from a mixture of the second torsional and second flexural modes. In Fig. 6(b), the vibrational mixed mode pattern that was recorded for this reed is shown. In Table I, all resonance frequencies that were

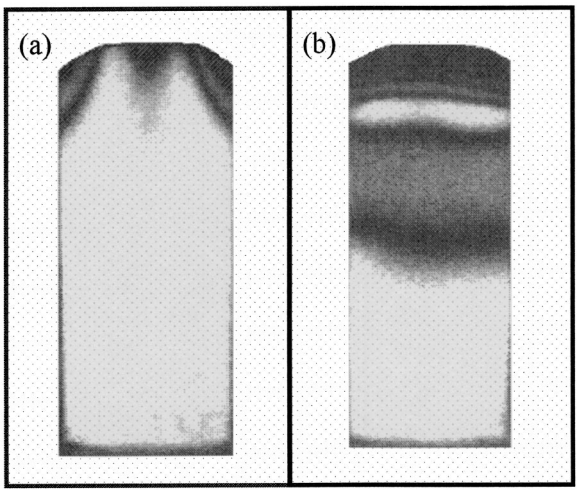


FIG. 5. Typical vibration patterns obtained under dry conditions (a) corresponding to the second torsional mode of a poor reed and (b) corresponding to the second flexural mode of a very poor reed of our sample set.

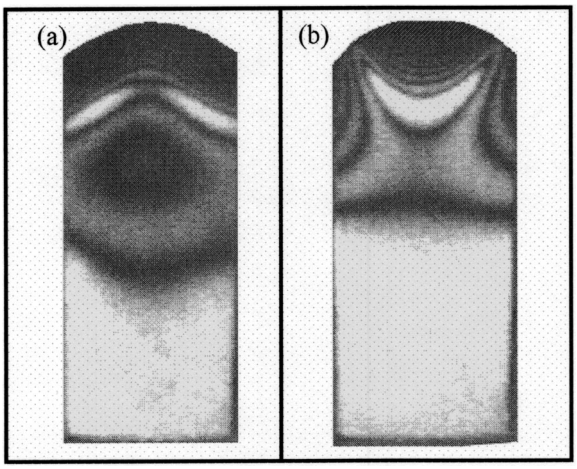


FIG. 6. Typical vibration patterns obtained under dry conditions for a mixed mode resulting from a combination of a second flexural and second torsional vibrational movement (a) for a good reed and (b) for a very good reed of our sample set.

detected under dry conditions are tabulated for all reeds ordered by musical quality categories. In addition, in order to indicate the resonance quality for each of the vibrational modes for all reeds, three different fonts are used: italic, normal, and bold type representing ''hardly detectable,'' ''good,'' and ''very strong'' resonances, respectively.

This first series of experiments under dry reed conditions suggests that the ''musical'' quality of a clarinet reed should be closely related to two different characteristics in its vibrational behavior that can quite easily be assessed experimentally: the first and for sure the most important criterion is the existence and the symmetry of the first torsional mode that is highly sensitive to the longitudinal symmetry of the reed for both geometric and mechanical characteristics of the raw material. A second, but less critical quality measure is the proximity of the natural resonance frequencies corresponding to the second torsional and the second flexural modes—a proximity that might produce mixed vibrational modes for the very best quality reeds.

B. Wet reed conditions

To adapt the experimental setup to wet reed conditions, a machine producing cold water vapor using ultrasonic microvalves was employed. After several tests, we concluded that this device is sufficiently efficient to stabilize the level of reed wetness for the natural frequency search, but not for the image recording; i.e., during the exposure time, the reeds changed their geometrical volume so much due to water-induced swelling or contracting that the mode patterns observed in the previous section became superposed by one or two large interference fringes roughly following the shape of the reeds. Since this volume change of the reeds results in a motion with more or less constant velocity during the exposure time, those fringes exhibit an intensity distribution that can be described by a sinc (sin(x)/x) function, rather than a Bessel function.[4] To eliminate the additional fringes, the AGFA holographic plates were replaced by Kodak SO-253 plates. While the Kodak plates provide a considerably lower quality in picture definition, they only require an exposure time of the order of 7 to 8 s instead of the 40 to 45 s typically needed with the high-resolution AGFA plates. This change was necessary to prevent the reed from drying or swelling too much during the exposure time, and resulted in fringe patterns very comparable to those shown above.

The most obvious change occurring in the vibrational behavior under wet conditions was a significant shift of all natural resonance frequencies toward lower values. This observation is readily explained by the fact that the infiltration of water in the reed material induces both a decrease in reed stiffness and an increase in specific gravity: as expected, the wetter the reeds, the lower their resonance frequencies. However, just how wet does a reed have to be to mimic playing conditions exactly? To answer this question in spite of its apparent complexity, we adopted the following protocol: first, we selected four *typical* reeds out of our 24-reed sample set of above. Two of those reeds were randomly chosen out of the very poor and the poor musical quality categories. The two others corresponded to the best reed of the good and the best reed of the very good musical quality category; i.e., the last ones listed in Table I for each of those categories. Then, one of the professional musicians joined us in the holography laboratory. The purpose of this very close collaboration was to determine the natural reed frequencies under real playing conditions precisely. To this end, the clarinet player intensively played each of the four reeds to achieve realistic, wet playing conditions, and we experimentally determined their natural frequencies immediately thereafter. In Table II, the resulting resonance frequencies for each observed vibrational mode are presented. The relative frequency shifts suffered by each of those reeds compared to the corresponding values measured under dry conditions just before the musician played them are also shown in Table II.

TABLE II. First four resonance frequencies determined for four typical reeds selected from the four different musical quality categories as classified by two professional musicians (see the text for the employed selection criteria). The cited frequencies were now measured under wet reed conditions (see the text for details). $\Delta\nu_2$ is the measured frequency difference between the second flexural and the second torsional mode. In parentheses, we give the percentage by that the resonance frequency decreased due to being intensively played by one of the musicians (see the text).

	1st flexural	1st torsional	2nd flexural	2nd torsional	$\Delta\nu_2$
Very poor	1890 Hz (−19%)		3380 Hz (−42%)	4000 Hz (−36%)	620 Hz
Poor	1760 Hz (−20%)		4200 Hz (−28%)	4790 Hz (−25%)	590 Hz
Good	1530 Hz (−32%)	2200 Hz (−37%)	3780 Hz (−36%)	4200 Hz (−29%)	420 Hz
Very good	1700 Hz (−27%)	2215 Hz (−39%)	3510 Hz (−44%)	3800 Hz (−39%)	290 Hz

Once these values were known, it was possible to reproduce realistic playing conditions by ourselves at any time, which allowed us to perform an in-depth study of the corresponding vibration patterns under wet reed conditions. The most striking result is a significant down shift of up to 44% in the measured resonance frequencies (see Table II). The new results concerning the vibration patterns, however, are a direct confirmation of the above conclusions regarding the relationship between vibrational behavior and musical quality: As before, under dry reed conditions, it was impossible for us to excite the first torsional mode for the reeds of the poor and the very poor quality categories, while this mode was always easily excited for the reeds of the good and the very good musical quality categories. Furthermore, the symmetry of the corresponding mode pattern appeared again to increase with reed quality. Differently than under dry conditions, however, the second torsional and the second flexural mode were now always well resolved even for the good and the very good reeds, which resulted in pure vibration patterns as those shown in Fig. 5. Consequently, no more mixed modes were found under wet conditions in our present sample set. However, the last two modes were still much closer in frequency for the very good reed than for all the others (see Table II). The disappearing of the mixed mode patterns under wet conditions can probably be understood in two different ways: either by a frequency shift that is considerably different for the two pure modes due to clearly different mechanical symmetry properties for the longitudinal and transversal reed modes, or by an important change in the mechanical coupling between those two modes under wet conditions.

IV. CONCLUSIONS

First, in good agreement with ''common sense,'' our study unambiguously demonstrates that a clear relationship between the vibrational behavior of clarinet reeds and their ''musical'' quality exists. In consequence, the clarinet reed quality can relatively easily be assessed experimentally using holography or any other experimental technique that allows observing the first three or four vibrational modes and their patterns.

Second, our results clearly show that the sound quality of a clarinet reed is directly dependant on the existence and symmetry of a ''strong'' first torsional mode. A second, but less important criterion to judge the musical quality from the vibrational behavior is the frequency interval between the second torsional and the second flexural mode. For the very best of our reeds, those two frequencies occurred in such a narrow range that the corresponding modes could be simultaneously excited, resulting in the formation of mixed mode patterns. It was further shown that the principal conclusion regarding the importance of the first torsional mode proves to be true for both dry and wet reed conditions. Consequently, it appears preferable to perform holographic quality tests under dry conditions.

We would like to point out that our main conclusion seems to be in excellent qualitative agreement with the empirical observation in many professional orchestras. We showed that the musical quality of clarinet reeds is closely related to the first torsional mode; i.e., its second vibrational mode. Our conclusion is reflected by the clarinetist's intuitive care to protect the corners of his reeds since he knows by experience how important those parts are for the musical quality of his instrument. In accordance, we showed that those corners are actually the vibrating parts of the reeds for the first torsional mode. This observation is in clear contrast to the musician's experience with double reeds, where the corners can even be slightly cut or damaged without any major deterioration of the musical quality.[5]

Tentatively, we would like to discuss the physical significance of our conclusions in the following. First, we would like to address the striking difference between the importance of the first flexural and the first torsional mode of the clarinet reed for its resulting musical quality. To this end, it should be remembered that the reed plays the role of a valve controlling the airflow into the instrument. As such, the movement related to the first torsional mode seems to be more efficient than the one of the first flexural one, since the former is located closer to the free tip of the reed inside of the player's mouth (see Figs. 3 and 4). Consequently, the first torsional mode is much less constrained by the pressure of the player's lips than the first flexural mode, allowing it to vibrate nearly without any restraints even under real playing conditions. In addition, the mass of the wooden part that has to be displaced during reed vibration is much lower for this mode due to the decreasing reed thickness toward the tip, which makes a faster reed response possible. This conclusion concerning a faster response dynamics is further supported by our observation that the curvature with which the reed bends away from the mouthpiece (defined as number of interference fringes per millimeter), is about 50% higher for the first torsional mode than for the first flexural mode [see Figs. 3 and 4(a)], which again allows a more efficient control of the airflow.

The importance of symmetry of the first torsional mode for high-quality sound production is demonstrated above by our experiments. This conclusion can again be readily understood in terms of efficient airflow control: The more the first torsional mode appears symmetric, the higher the airflow can be for a given constraint by the player's lips [compare Figs. 4(a) and (b)]; i.e., the more this mode is symmetric, the more the air can pass by both reed corners into the clarinet, allowing a faster instrument response as witnessed by our two professional musicians.

Finally, we have to address the delicate question of how the second flexural and torsional mode can have any influence on the musical quality considering that their frequencies are so much higher than the dominant spectrum of musical tones around or below 2 kHz (see Tables I and II). The mixed modes shown in Fig. 6 are only observed for good and very good reeds. Whenever we witnessed the existence of those mixed modes, they were particularly symmetric as those shown in Figs. 2 and 6. In this sense, the presence of high-frequency mixed modes could simply be an additional and extremely sensitive measure for the reed's vibrational symmetry that we already discussed above. Alternatively, the existence of those high-frequency modes could favor the responsiveness of the reed as perceived by the players with

regard to its transient behavior. In this manner, the high-frequency modes could improve the attack response of the reed and, thus, participate in the creation of the ''woody'' timbre characteristic for the clarinet. Finally, we would also like to point out that the measured frequencies of the second-order modes are roughly three times higher than the one of the first flexural mode. Consequently, we would like to propose that those high-frequency modes might possibly help to emphasize the second of the well-known odd-numbered harmonics of the clarinet.[6] This interpretation is, of course, very tentative and is actually the subject of further investigations.

As a final point, we hope that reed manufacturers might be stimulated by our present study to produce ''high-quality'' reeds—beside their standard ones—that have been tested for the existence of the first torsional vibrational mode under relatively mild excitation conditions. Such a test does not necessarily have to be performed by holography, but could be automated using a more simple experimental technique such as speckle interferometry, for instance. In addition, reed designers might be inspired by our present findings to accomplish reeds that favor an easy excitation of the first torsional vibrational mode.

ACKNOWLEDGMENTS

The holographic measurements were carried out in the Centre de Travaux Expérimentaux of the Physics Department in Ecole Polytechnique. The authors wish to thank Claude Chesnais-Popovic and Daniel Paget for their hospitality and their support, and the freedom they offered us to carry out this study. We would also like to acknowledge the stimulating discussions with Vincent Gibiat and Xavier Boutillon at the very early stages of the present study.

[1] S. C. Thompson, ''The effect of the reed resonance on woodwind tone production,'' J. Acoust. Soc. Am. **66**, 1299–1307 (1979).
[2] P. L. Hoekje and G. M. Roberts, ''Observed vibration patterns of clarinet reeds,'' J. Acoust. Soc. Am. **99**, 2462(A) (1996).
[3] I. M. Lindevald and J. Gower, ''Vibrational modes of clarinet reeds,'' J. Acoust. Soc. Am. **102**, 3085 (A) (1997).
[4] Y. I. Ostrovsky, M. M. Butusov, and G. V. Ostrovskaya, *Interferometry by Holography* (Springer, Berlin, 1980).
[5] J. Veragh and A. Leroy, private communication.
[6] T. D. Rossing, *The Science of Sound* (Addison-Wesley, New York, 1990).

Theoretical calculation and experimental study on the third-order nonlinearity parameter C/A for organic liquids and biological fluids

Xiao-chen Xu, Feng Mao,[a] Xiu-fen Gong,[b] and Dong Zhang
State Key Lab of Modern Acoustics, Institute of Acoustics, Nanjing University, Nanjing 210093, People's Republic of China

(Received 8 February 2002; accepted for publication 8 December 2002)

In the present paper we deal with calculating and measuring the third-order nonlinearity parameter C/A for organic liquids and biological fluids. For the organic liquids, the second-order derivative of sound speed with respect to pressure at constant entropy is discussed in terms of thermodynamic parameters, and a formula for calculating C/A is established. Calculated results of C/A for various organic liquids demonstrate that omitting the second-order derivative of sound speed with respect to sound pressure in an isentropic process, could overestimate the value of C/A about 20%–30%. For measuring C/A, the finite-amplitude insert substitution technique is employed. A theoretical description of the sound pressure amplitude for the third-order harmonics is obtained using the perturbation method, in which the sound pressure amplitude of the third-order harmonics is shown approximately as a parabolic function of the propagation distance in the medium. By measuring and comparing the sound pressure amplitudes of the third-order harmonics when inserting a test sample, the reference liquid, and a comparative liquid, respectively, the value of C/A for the sample liquid is determined with reference to known C/A values of reference and comparative liquids. An experimental setup is developed, and the measured values of C/A for the organic liquids agree with the predicted ones calculated by the thermodynamic methods. Furthermore, the finite-amplitude method is extended to measure the C/A parameter for biological fluids, and the measured values of C/A for several biological fluids are also presented. [DOI: 10.1121/1.1553460]

PACS numbers: 43.80.Cs, 43.80.Ev [FD]

I. INTRODUCTION

Exploring new acoustic imaging parameter is one of the research focuses in biomedical ultrasound at the present time.[1,2] Significant progress on the study of the second-order nonlinearity parameter B/A have been made over the past decade. It has been demonstrated that the values of B/A are strongly influenced by the compositions and the structural features of biological tissues.[3–6] Furthermore, much attention has been attracted on the nonlinearity parameter B/A imaging, which could be have a potential and significant application in medical diagnosis.[7–9] Recently, it is reported that higher harmonics ($n \geqslant 3$) possess superior characteristics to second harmonics. The higher harmonic images of a phantom showed exceptionally improve clarity and sharper contrast between the different structures of the phantom.[10] The third-order nonlinearity parameter C/A is a measure of the third-order harmonics in a fluid or fluid-like medium. Therefore, it is very interesting to study the third-order nonlinearity parameter C/A not only for its academic meaning, but also for its prospective application.

Comparing with the second-order nonlinearity parameter B/A, fewer works have been done concerning the third-order nonlinearity parameter C/A. Coppens *et al.* gave a theoretical expression of C/A in terms of B/A and the second-order derivative of sound speed with respect to pressure.[11] Hagelberg found that the second derivative term may be obtained from the equations of sound speed versus pressure for a given temperature, and values of C/A for water and 1-propanol were estimated.[12,13] Blackstock studied the propagation of plane sound waves of finite amplitude in a lossless fluid, and gave an solution for the third-order harmonics by power series.[14]

Our purpose in the present paper is to investigate further this parameter in theory as well as in measurement. On the base of van Der Waals' equation, Schaaffs derived an expression of sound velocity for an organic liquid in terms of thermodynamic parameters relating with molecular weight, molecular volume, and density. By using this expression, we calculate the second-order derivative of sound speed with respect to pressure in the formula of C/A obtained by Coppens.[11] Then a theoretical formula is developed for calculating C/A of the organic liquids, values of C/A for several kinds of organic liquids are calculated, and the influence of the second-order derivative term on the value of C/A is estimated.

However, for the experimental determination of C/A values, the thermodynamic method is much complicated and is not suitable for biological fluids and tissues. A finite-amplitude method is developed to measure the value of C/A,

[a]Present address: Department of Physics, HuBei Sanxia University, Yichang 443001, People's Republic of China.
[b]Author to whom correspondence should be addressed. Electronic mail: zligong@nju.edu.cn

which is easier to be carried out in the laboratory. Solving fluid-dynamic equations in third-order approximations with the perturbation method, we derive an expression for the sound pressure amplitude of the third-order harmonics, and employ it as the principle in the measurement. In order to avoid from measuring the absolute value of the sound pressure, the finite-amplitude insert substitution technique[18,19] is applied. By measuring and comparing the sound pressure amplitudes of third harmonics with inserting the reference liquid, a comparative liquid, and the sample, respectively, the value of C/A for the sample liquid is determined under the circumstances of the known values C/A for the reference and the comparative liquids. Values of C/A for several organic liquids are measured using the finite-amplitude method, and compared with the calculated values by thermodynamic method. Finally, this finite-amplitude method is extended to measure the values of C/A for biological fluids.

II. CALCULATION ON C/A FOR ORGANIC LIQUIDS

A. C/A expression for organic liquids using thermodynamics method

When a finite-amplitude wave propagates in a homogeneous medium, the nonlinear effects occurred and the second-order and the third-order harmonics will be generated. The nonlinearity parameters B/A and C/A arise from the Taylor series expansion expressing the variations in pressure with density in the medium. It represents the fact that density does not follow linearly the changes in applied pressure, and may be put in the form

$$p-p_0=A\left(\frac{\rho-\rho_0}{\rho_0}\right)+\frac{B}{2}\left(\frac{\rho-\rho_0}{\rho_0}\right)^2+\frac{C}{6}\left(\frac{\rho-\rho_0}{\rho_0}\right)^3+\cdots, \tag{1}$$

where p is the pressure in the medium and ρ is the density, p_0, ρ_0 are the static values of the same parameters. Coefficients $A=\rho_0(\partial p/\partial\rho)_{s,0}=\rho_0 c_0^2$, $B=\rho_0^2(\partial^2 p/\partial\rho^2)_{s,0}$, and $C=\rho_0^3(\partial^3 p/\partial\rho^3)_{s,0}$. All partial derivations are evaluated at equilibrium configuration and at constant entropy, which are denoted by subscripts "o" and "s," c is the sound speed. The parameters of B/A and C/A are the ratio of the quadratic to linear terms and of the cubic to linear terms in the Taylor series, respectively,

$$\frac{B}{A}=2\rho_0 c_0\left(\frac{\partial c}{\partial p}\right)_{s,0}, \tag{2}$$

$$\frac{C}{A}=\frac{\rho_0^2}{c_0^2}\left(\frac{\partial^3 p}{\partial\rho^3}\right)_{s,0}. \tag{3}$$

To simplify the expression in the following, the subscript "0" is omitted, but all the variables take the values at the equilibrium state. Upon expanding the derivative $(\partial^3 p/\partial\rho^3)_s$, and using the thermodynamic relations, Coppens obtained the formula of C/A as follows:[11]

$$\frac{C}{A}=\frac{3}{2}\left(\frac{B}{A}\right)^2+2\rho^2 c^3\left(\frac{\partial^2 c}{\partial p^2}\right)_s. \tag{4}$$

The key point of calculating C/A is how to determine the value of $(\partial^2 c/\partial p^2)_s$. Since the B/A parameter of many organic liquids have been extensively studied, and their values can be found in the literature. In 1939, Schaaffs found that the van Der Waals' equation could be applied to all organic liquids at room temperature and under atmospheric pressure, and the sound velocity for organic liquids could be expressed as[15]

$$c=\left\{\gamma RT\left[\frac{M}{3(M-b\rho)^2}-\frac{2}{(M-b\rho)}\right]\right\}^{1/2}, \tag{5}$$

where γ is the ratio of specific heats, M is the molecular weight, T is the absolute temperature, R is the universal gas constant and b is the corrections in Van Der Waals' equation for considering the actual volume of molecular. Based on this formula of Schaaffs, we calculate the second term of Eq. (4) in order to express it by thermodynamic parameters. Generally, b equals fourfold actual molecular volume in a molar organic liquid, and[16,17]

$$\left(\frac{\partial b}{\partial\rho}\right)_T=-\frac{2M}{3\rho^2},\quad \left(\frac{\partial b}{\partial T}\right)_\rho=0. \tag{6}$$

Differentiating Eq. (5) with respect to pressure and using Eq. (6), we obtain

$$\left(\frac{\partial c}{\partial p}\right)_s=\left(1-\frac{1}{\gamma}\right)\frac{ck}{2T\alpha}+\frac{1}{2\rho c}J(\xi), \tag{7}$$

where

$$\xi=\frac{M}{b\rho},\quad k=-\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_T,\quad \alpha=\frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_p,$$

$$J(\xi)=\frac{2(2\xi-3)^2}{3(\xi-1)(6-5\xi)}.$$

It is not difficult to deduce the expression of $(\partial^2 c/\partial p^2)_s$ from Eq. (7). In the case of an isentropic state, $(\partial\rho/\partial p)_s=c^{-2}$, $(\partial\xi/\partial p)_s=\xi(3-2\xi)/(3\rho c^2)$, and $(\partial T/\partial p)_s=(\gamma-1)k/\gamma a$, we differentiate Eq. (5) with respect to pressure again, and obtain

$$\left(\frac{\partial^2 c}{\partial p^2}\right)_s=\frac{c^2k\rho-1}{2Tc\rho k\alpha}\left(\frac{\partial k}{\partial p}\right)_s-\frac{c^2k\rho-1}{2Tc\rho\alpha^2}\left(\frac{\partial\alpha}{\partial p}\right)_s-\frac{(c^2k\rho-1)^2}{4T^2c^3\rho^2\alpha^2}-\frac{\theta^2(3+2\theta)}{3\rho^2c^3(\theta-3)^2}, \tag{8}$$

where $\theta=(3M/RTk\rho+9)^{1/2}$. For most organic liquids, $[(\partial k/\partial p)_s/k]<0.1$, and α is independent of pressure at room temperature and under atmospheric pressure, i.e., $(\partial\alpha/\partial p)_s=0$.[20–22] Neglecting the first and second terms on the right-hand side of this equation, we reduce Eq. (8) to the form

$$\left(\frac{\partial^2 c}{\partial p^2}\right)_s=-\frac{(c^2k\rho-1)^2}{4T^2c^3\rho^2\alpha^2}-\frac{\theta^2(3+2\theta)}{3\rho^2c^3(\theta-3)^2}. \tag{9}$$

Substituting Eq. (9) into Eq. (4) yields

$$\left(\frac{C}{A}\right)=\frac{3}{2}\left(\frac{B}{A}\right)^2-\left[\frac{(c^2k\rho-1)^2}{2T^2\alpha^2}+\frac{2\theta^2(3+2\theta)}{3(\theta-3)^2}\right]. \tag{10}$$

TABLE I. Values of B/A, sound velocity, thermodynamic parameters, and calculated C/A using the thermodynamic method for various organic liquids (25 °C).

Organic liquids	B/A	c_0 (m s^{-1})	ρ_0 (kg m^{-3})	$k\times10^{11}$ (m^2 N^{-1})	$\alpha\times10^5$ (K^{-1})	$M\times10^3$ (kg mol^{-1})	$(C/A)'$	$(C/A)''$	C/A
Ethanol	10.57 (Ref. 11)	1180 (Ref. 24)	789 (Ref. 24)	119.1 (Ref. 21)	104 (Ref. 21)	46	167.59	−32.58	135.01
Methanol	9.47 (Ref. 23)	1123 (Ref. 24)	792 (Ref. 24)	129.3 (Ref. 21)	126 (Ref. 21)	32	134.52	−35.56	98.96
Carbon tetrachloride	9.62 (Ref. 22)	928.7 (Ref. 24)	1595 (Ref. 24)	90.4 (Ref. 21)	123.6 (Ref. 21)	154	138.82	−32.08	106.74
Benzene	9.03 (Ref. 11)	1326 (Ref. 24)	878 (Ref. 24)	94.6 (Ref. 21)	123.7 (Ref. 21)	78	122.31	−32.32	89.99
Glycol	9.58 (Ref. 24)	1683.5 (Ref. 24)	1113 (Ref. 24)	37.2 (Ref. 21)	63.8 (Ref. 21)	62	137.66	−33.72	103.94
n-hexane	9.47 (Ref. 24)	1103 (Ref. 24)	654 (Ref. 24)	156 (Ref. 21)	135 (Ref. 21)	86	134.52	−31.55	102.97
Pentane	9.72 (Ref. 11)	1008 (Ref. 24)	626 (Ref. 24)	238 (Ref. 21)	161 (Ref. 21)	72	141.72	−31.55	110.17
Heptane	11.14 (Ref. 25)	1162 (Ref. 24)	684 (Ref. 24)	132 (Ref. 21)	124 (Ref. 21)	100	186.15	−32.15	154.00
Octane	11.34 (Ref. 25)	1197 (Ref. 24)	703 (Ref. 24)	119 (Ref. 21)	116 (Ref. 21)	126	192.89	−33.45	159.44
Nonane	11.27 (Ref. 25)	1248 (Ref. 24)	738 (Ref. 24)	123.1 (Ref. 21)	107 (Ref. 21)	128	190.52	−33.83	156.69
n-propyl alcohol	10.46 (Ref. 11)	1223 (Ref. 24)	804 (Ref. 24)	89.5 (Ref. 21)	95.6 (Ref. 21)	60	164.12	−31.39	132.73
n-butyl alcohol	10.72 (Ref. 11)	1268 (Ref. 24)	810 (Ref. 24)	88.84 (Ref. 21)	95 (Ref. 21)	74	172.38	−31.91	140.47
Cyclohexane	10.1 (Ref. 11)	1284 (Ref. 24)	779 (Ref. 24)	111 (Ref. 21)	119 (Ref. 21)	86	153.02	−32.39	120.63

In order to discuss conveniently, rewrite Eq. (10) as

$$\left(\frac{C}{A}\right)=\left(\frac{C}{A}\right)'+\left(\frac{C}{A}\right)'', \tag{11}$$

where

$$\left(\frac{C}{A}\right)'=\frac{3}{2}\left(\frac{B}{A}\right)^2, \tag{12}$$

$$\left(\frac{C}{A}\right)''=-\left[\frac{(c^2k\rho-1)^2}{2T^2\alpha^2}+\frac{2\theta^2(3+2\theta)}{3(\theta-3)^2}\right]. \tag{13}$$

B. Numerical calculation of *C/A* for organic liquids

In the present paper, by using the known values of B/A, the sound velocity and the thermodynamic parameters the values of C/A for several organic liquids are calculated from Eqs. (11), (12), and (13), which are listed in Table I. The calculated values of (C/A), $(C/A)'$, $(C/A)''$ are also listed in this table.

III. DETERMINATION OF *C/A* USING FINITE-AMPLITUDE METHOD

In this section, we discuss the principle of measuring the third-order nonlinearity parameter C/A using the finite-amplitude method, which is based on the accumulation solution for the sound pressure of the third-order harmonics in lossless fluids.[26]

Suppose that a finite-amplitude plane wave propagates in a lossless fluid, the set of fluid-dynamic equations is

$$\frac{\partial\rho}{\partial t}+u\frac{\partial\rho}{\partial x}+\rho\frac{\partial u}{\partial x}=0, \tag{14a}$$

$$\rho\frac{\partial u}{\partial t}+\rho u\frac{\partial u}{\partial x}=-\frac{\partial p}{\partial x}, \tag{14b}$$

$$p=p(\rho,s). \tag{14c}$$

We use the method of perturbation to solve Eq. (14), expressing ρ, u, p as a series,

$$\rho=\sum_{n=0}^{\infty}\rho^{(n)},\quad u=\sum_{n=1}^{\infty}u^{(n)},\quad p=\sum_{n=1}^{\infty}p^{(n)}, \tag{15}$$

and substituting these series into Eq. (14), then, the equations in the second- and the third-order approximations can be expressed as follows:

$$\begin{aligned}\frac{\partial^2p^{(2)}}{\partial x^2}-\frac{1}{c_0^2}\frac{\partial^2p^{(2)}}{\partial t^2}&=-\frac{\partial}{\partial x}\left\{\rho_0^{-1}c_0^2(\beta-1)\frac{\partial[\rho^{(1)}]^2}{\partial x}+\rho^{(1)}\frac{\partial u^{(1)}}{\partial t}\right.\\&\left.+\rho_0u^{(1)}\frac{\partial u^{(1)}}{\partial x}\right\}+\rho_0^{-1}c_0^2(\beta-1)\frac{\partial^2[\rho^{(1)}]^2}{\partial x^2}\\&+\frac{\partial^2\rho^{(1)}u^{(1)}}{\partial x\partial t}-\rho_0^{-1}c_0^2(\rho-1)\frac{\partial^2[\rho^{(1)}]^2}{\partial t^2},\end{aligned} \tag{16}$$

$$\begin{aligned}\frac{\partial^2p^{(3)}}{\partial x^2}-\frac{1}{c_0^2}\frac{\partial^2p^{(3)}}{\partial t^2}&=-2\rho_0^{-1}(\beta-1)\frac{\partial^2}{\partial t^2}[\rho^{(1)}\rho^{(2)}]-\rho_0^{-2}(\eta-1)\frac{\partial^2}{\partial t^2}[\rho^{(1)}]^3\\&-\frac{\partial}{\partial x}\left[\rho^{(1)}\frac{\partial u^{(2)}}{\partial t}+\rho^{(2)}\frac{\partial u^{(1)}}{\partial t}+\rho_0u^{(1)}\frac{\partial u^{(2)}}{\partial x}\right.\\&\left.+\rho^{(1)}u^{(1)}\frac{\partial u^{(1)}}{\partial x}+\rho_0u^{(2)}\frac{\partial u^{(1)}}{\partial x}\right]+\frac{\partial^2}{\partial t\partial x}[\rho^{(1)}u^{(2)}\\&+\rho^{(2)}u^{(1)}].\end{aligned} \tag{17}$$

Assuming an initially sinusoidal motion at the boundary position $x=0$, from Eq. (16) we obtain

$$\rho^{(1)}=c_0^{-2}p_0e^{j(\omega t-kx)}, \tag{18a}$$

$$\rho^{(2)}=\frac{1}{2\rho_0c_0^5}\beta\omega p_0^2xe^{2j(\omega t-kx)}, \tag{18b}$$

$$u^{(1)}=\frac{1}{\rho_0c_0}p_0e^{j(\omega t-kx)}, \tag{18c}$$

$$u^{(2)}=\frac{1}{2\rho_0^2c_0^4}\beta\omega p_0^2xe^{2j(\omega t-kx)}, \tag{18d}$$

$$p^{(1)}=p_0e^{j(\omega t-kx)}, \tag{18e}$$

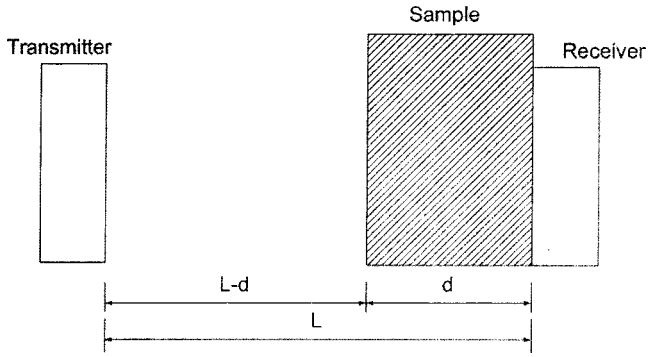


FIG. 1. Schematic diagram of the transducers and the sample.

$$p^{(2)}=\frac{1}{2\rho_0 c_0^3}\beta\omega p_0^2 x e^{2j(\omega t-kx)}, \tag{18f}$$

where p_0 is the amplitude of the fundamental sound pressure, ω is the angular frequency, and $\beta=1+B/2A$ is the second-order nonlinearity coefficient. Substituting Eq. (18) into the right side of Eq. (17), we find that the amplitude of the third-order harmonics is

$$p_3=Gfp_0^3(Ff\beta^2x^2+2\eta x), \tag{19}$$

where $G=\pi/(2\rho_0^2c_0^5)$, $F=6\pi/c_0$, and $\eta=1+C/2A$.

IV. EXPERIMENT

A. Measurement principle

Figure 1 shows the schematic diagram of the transducers and the sample. The transmitter is excited at the fundamental frequency f and emits ultrasonic rf pulses into the reference medium (water). The receiver detects ultrasonic rf pulses at the third-order harmonic frequency $3f$. In the measurement, the finite-amplitude insert substitution technique is used, which can avoid the measurement of the absolute value of the sound pressure.[18,19] From Eq. (19), the amplitude of the third harmonic in the reference liquid without insertion of a sample is

$$p_{3w}=G_wfp_0^3(F_wf\beta_w^2L^2+2\eta_wL), \tag{20}$$

where $G_w=\pi/(2\rho_w^2c_w^5)$, $F_w=6\pi/c_w$, $\eta_w=1+(C/A)_w/2$, and L is the distance between the transmitter and the receiver.

Inserting a sample with thickness d between the transmitter and the receiver, the pressure amplitude of the third-order harmonic is

$$p_{3s}=D_{1s}D_{2s}G_wfp_0^3[F_w\beta_w^2(L-d)^2+2\eta_w(L-d)]+D_{2s}^3D_{1s}G_s(F_s\beta_s^2d^2+2\eta_sd), \tag{21}$$

where the subscript "s" denotes the sample, therefore, the definitions of F_s, β_s, and η_s for the sample are corresponding to those of F_w, β_w, and η_w for water, except that the subscript "w" is substituted by "s." D_{1s}, D_{2s} are the transmission coefficients for sound pressure from the reference medium to the sample and from the sample to the reference medium, and defined by

$$D_{1s}=\frac{2(\rho c)_s}{(\rho c)_w+(\rho c)_s}, \tag{22a}$$

$$D_{2s}=\frac{2(\rho c)_w}{(\rho c)_w+(\rho c)_s} \tag{22b}$$

From Eqs. (20) and (21), we obtain

$$\frac{p_{3s}}{p_{3w}}=D_{1s}D_{2s}\frac{F_wf\beta_w^2(L-d)^2+2\eta_w(L-d)}{F_wf\beta_w^2L^2+2\eta_wL}+D_{1s}D_{2s}^3\frac{G_s}{G_w}\frac{F_sf\beta_s^2d^2+2\eta_sd}{F_wf\beta_w^2L^2+2\eta_wL}. \tag{23}$$

If a comparative liquid with known value C/A substitutes the sample liquid, a similar expression is then given by

$$\frac{p_{3c}}{p_{3w}}=D_{1c}D_{2c}\frac{F_wf\beta_w^2(L-d)^2+2\eta_w(L-d)}{F_wf\beta_w^2L^2+2\eta_wL}+D_{1c}D_{2c}^3\frac{G_c}{G_w}\frac{F_cf\beta_c^2d^2+2\eta_cd}{F_wf\beta_w^2L^2+2\eta_wL}, \tag{24}$$

where p_{3c}, F_c, β_c, η_c, G_c, D_{1c}, D_{2c} are corresponding to the definitions of p_{3s}, F_s, β_s, η_s, G_s, D_{1s}, D_{2s}, but the subscript "s" is replaced with "c," which stands for the comparative liquid. Using Eqs. (23) and (24), we obtain the expression for $(C/A)_s$ as

$$\left(\frac{C}{A}\right)_s=\frac{E_1E_3}{E_2}+E_4, \tag{25}$$

where

$$E_1=\frac{G_w}{D_{1s}D_{2s}^3dG_s}\left[\frac{p_{3s}}{p_{3w}}\frac{L^2}{c_w}\beta_w^2-D_{1s}D_{2s}\frac{\beta_w^2}{c_w}(L-d)^2-D_{1s}D_{2s}^3\frac{G_s}{G_wc_s}\beta_s^2d^2\right],$$

$$E_2=\frac{p_{3c}}{p_{3w}}\frac{L^2}{c_w}\beta_w^2-D_{1c}D_{2c}\frac{\beta_w^2}{c_w}(L-d)^2-D_{1c}D_{2c}^3\frac{G_c}{G_wc_c}\beta_c^2d^2,$$

$$E_3=2\left[D_{1c}D_{2c}\eta_w(L-d)+D_{1c}D_{2c}^3\frac{G_c}{G_w}\eta_cd-\frac{p_{3c}}{p_{3w}}\eta_wL\right],$$

$$E_4=\frac{2G_w}{D_{1s}D_{2s}^2dG_s}\left[\frac{p_{3s}}{p_{3w}}\eta_wL-D_{1s}D_{2s}\eta_w(L-d)\right]-2.$$

In Eq. (25), we find that the value of C/A is independent on the source amplitude and the fundamental frequency. After measuring the p_{3w}, p_{3s}, and p_{3c}, the third-order nonlinearity parameter $(C/A)_s$ of the sample liquid can be calculated with Eq. (25).

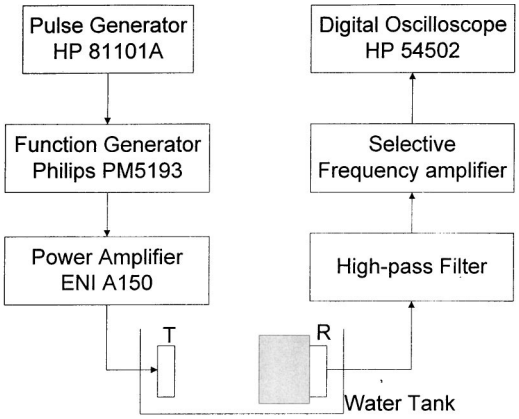


FIG. 2. Arrangement of the experimental setup for measuring C/A with the finite-amplitude method.

B. Experiment and results

1. Experimental system

An arrangement of the experimental setup is shown in Fig. 2. The measurements are performed in a water tank. The transmitter, a plane PZT transducer with 1.5 cm radius, has a fundamental frequency of 1 MHz. A function generator (Philips PM5193) operating in a tone burst mode with a 1 KHz frequency repetition rate, is connected to a power amplifier (ENI A 150). A plane PZT transducer with 1 cm radius and resonance frequency 3 MHz is used as the receiver. After passing through a high-pass filter and a selective frequency amplifier, the amplitude of the third-order harmonics is recorded by a digital oscilloscope (HP 54502). All measurements are carried out at 25 °C.

2. Experiment results

The dependence of the axial sound pressure amplitudes for the second-order and the third-order harmonics in water on the distance from the source are measured and shown in Fig. 3 as dot points. For comparison, the calculated corresponding curves using Eq. (22) are also given in this figure as solid lines, in which $\rho_w = 1000\,\mathrm{kg/m^3}$, $c_w = 1500\,\mathrm{m/s}$, $(B/A)_w = 5.2$, and $(C/A)_w = 32$.[11] Results show that the measured curve agrees with the calculated data.

TABLE II. Comparisons of the predicted values of C/A using the thermodynamic method and the measured ones using the finite-amplitude method for various organic liquids.

Organic liquids	Predicted value	Measured value
Methanol	98.96	106.9
Carbon tetrachloride	106.74	115.5
Benzene	89.99	87.6
Glycol	103.94	107.1
Heptane	154.00	138.1
N-butyl alcohol	140.47	127.8

In measurements of C/A, the distance between the transmitter and the receiver is $L = 8$ cm, and a sound permeable sample box of thickness $d = 3$ cm for organic liquids or $d = 2$ cm for biological fluids, is inserted between the transmitter and the receiver, and located to be close to the receiver. Six kinds of organic liquids and four kinds of biological fluids are used as the sample liquids. The reference medium is distilled water with known $(C/A)_w = 32$,[11] and ethanol is used as the comparative medium with known $(C/A)_c = 135.01$, which is obtained from Eq. (11). Using the finite-amplitude insert substitution technique, we measure p_{3w}, p_{3c}, and p_{3s}, respectively. Then the values of C/A for samples are determined from Eq. (25) with the ratios of p_{3s}/p_{3w} and p_{3c}/p_{3w}. The results of C/A values for organic liquids shown in Table II agree well with the theoretical predictions by thermodynamic calculation. Furthermore, the finite-amplitude method is extended to measure the values of C/A for some biological fluids, the experimental results of which are presented in Table III.

V. DISCUSSION AND CONCLUSION

From Table I, we find that the values of $(C/A)''$ for organic liquids are far less than those of $(C/A)'$, and the values of $(C/A)''$ are negative. The results suggest that the value of C/A for the organic liquid is mainly dependent on

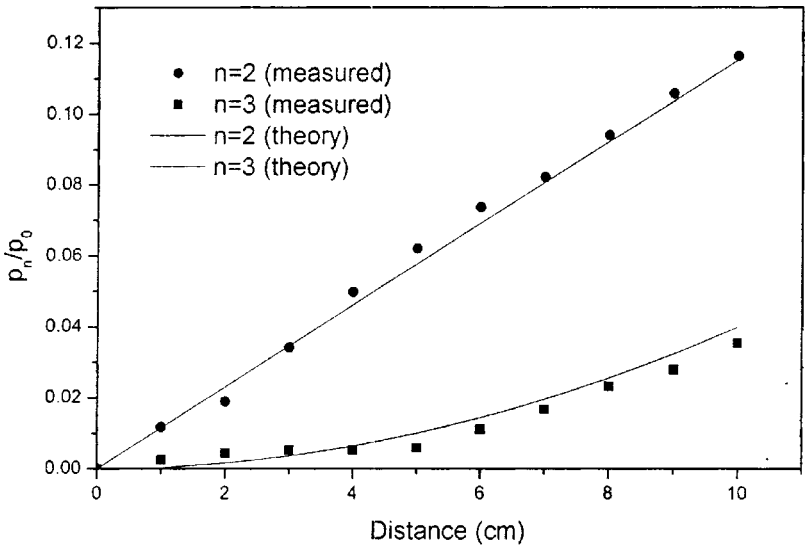


FIG. 3. Curves of axial sound pressure amplitude for the second-order and third-order harmonics in water versus distance from the source by theoretical calculation and measurement.

TABLE III. Measured values of C/A for various biological liquids using the finite-amplitude method.

Samples	B/A	Sound velocity (m/s)	Density (kg/m^3)	C/A
Porcine whole blood	5.8 (Ref.18)	1559	1053	42.1
Human whole blood	6.3	1532	1057	46.9
Dextrose (5%)	5.2 (Ref.18)	1518	1028	43.7
Dextrose (10%)	5.3 (Ref.18)	1530	1047	46.8

the value of $(C/A)'$, which agrees basically with the research results of Coppens *et al.*[11] They also proposed an approximate expression for C/A, i.e., $C/A=(3/2)(B/A)^2$, in which the effect of $(C/A)''$ was ignored. But the results presented here demonstrate that the ratio of $(C/A)''$ and $(C/A)'$ is about 20%–30%, and the value of C/A may be overestimated about 20%–30% without considering $(C/A)''$. On the other hand, we can find that the values of $(C/A)''$ for the organic liquids studied in this paper is in the range of -36–-30. Taking that this magnitude, on average, as "-33" is reasonable, consequently, the aforementioned expression for C/A may be modified as

$$\frac{C}{A}=\frac{3}{2}\left(\frac{B}{A}\right)^2-33. \tag{26}$$

Generally, the value of B/A for organic liquids is in the range of 8–12, which is obtained from the theoretical calculation or measurements. Therefore, the values of C/A of organic liquids may be in the range of 63–183 by means of Eq. (26).

In Table II, we can find that the measured values of C/A for organic liquids agree roughly with the calculated ones from the thermodynamic methods. In our measurement, the sound velocity is measured by using a pulse transmission method. The second-order nonlinearity parameter B/A is measured using the finite-amplitude insert substitution method, in which the absolute measurement of sound pressure is unnecessary, and the effects of diffraction could be effectively reduced. The values of p_{3s}/p_{3w}, p_{3c}/p_{3w} are measured with a precise digital oscilloscope. The main errors of the measurement may come from the measurements of velocity and second-order nonlinearity parameter B/A, and neglecting the sound attenuations of the sample. For the value of C/A, the accuracy associated with the quantities is estimated to be about 10%. Further development of this work with consideration of effects of sound attenuation and sound diffraction in the fluid dynamic equations may complete the finite-amplitude method to measure the third-order nonlinearity parameter C/A for a lossy medium, especially for biological tissues.

ACKNOWLEDGMENTS

The authors wish to thank Professor F. Dunn for his valuable comments in preparing the final version of this paper.

[1] X. F. Gong, "Research progress on acoustic nonlinearity in medical ultrasound," Prog. Phys. **16**, 286–298 (1996).
[2] F. A. Duck, "Nonlinear acoustics in diagnostic ultrasound," Ultrasound Med. Biol. **28**, 1–8 (2000).
[3] R. L. Errabolu, C. M. Sehgal, R. C. Bahn, and J. F. Greenleaf, "Measurement of ultrasonic nonlinear parameter in excised fat tissues," Ultrasound Med. Biol. **14**, 137–146 (1988).
[4] C. M. Sehgal, R. C. Bahn, and J. F. Greenleaf, "Measurement of the acoustic nonlinearity parameter B/A in human tissues by a thermodynamic method," J. Acoust. Soc. Am. **76**, 1023–1029 (1984).
[5] W. K. Law, L. A. Frizzell, and F. Dunn, "Determination of the nonlinearity parameter B/A of biological media," Ultrasound Med. Biol. **11**, 307–318 (1985).
[6] J. Zhang, M. Kuhlenschmidt, and F. Dunn, "Influences of structural factors of biological media on the acoustic nonlinearity parameter B/A," J. Acoust. Soc. Am. **89**, 80–91 (1991).
[7] N. Ichida, T. Sato, and M. Linzer, "Imaging the nonlinear ultrasonic parameter," Ultrason. Imaging **9**, 295–299 (1983).
[8] D. Zhang, X. F. Gong, and S. G. Ye, "Acoustic nonlinearity parameter tomography for biological specimens via measurement of the second harmonics," J. Acoust. Soc. Am. **99**, 2397–2402 (1996).
[9] D. Zhang and X. F. Gong, "Experimental investigation of acoustic nonlinearity parameter tomography for excised pathological biological tissues," Ultrasound Med. Biol. **25**, 593–599 (1999).
[10] A. Bouakaz and N. de. Jong, "Superharmonic tissue imaging," *Abstract for the 16th International Symposium on Nonlinear Acoustics*, 2002, p. 141.
[11] A. B. Coppens, R. T. Beyer, M. B. Seiden, J. Donohue, F. Guepin, R. Hodson, and C. Townsend, "Parameter of nonlinearity in fluids II," J. Acoust. Soc. Am. **38**, 797–804 (1965).
[12] M. P. Hagelberg, G. Holton, and S. Kao, "Calculation of B/A for water from measurements of ultrasonic velocity versus temperature and pressure to 10000," J. Acoust. Soc. Am. **41**, 564–567 (1967).
[13] M. P. Hagelberg, "Ultrasonic-velocity measurements and B/A for 1-propanol at pressure to 100000," J. Acoust. Soc. Am. **47**, 158–162 (1970).
[14] D. T. Blackstock, "Propagation of plane sound waves of finite amplitude in nondissipative fluids," J. Acoust. Soc. Am. **34**, 9–30 (1962).
[15] W. Schaaffs, "Zur bestimmung von molekülradien organischer fiussigkeiten aus schallgeschwindigkeit und dichte," Z. Phys. **114**, 110 (1939).
[16] W. Schaaffs, "Bemerküngen zur berechnung des molekulradius aus molvolumen und schallgeschwindigkeit," Z. Phys. **115**, 69 (1940).
[17] W. Schaaffs, "Über schallgeschwindigkeit und konstitution in flüssigen organischen verbindungen," Ann. Phys. (Leipzig) **40**, 393 (1941).
[18] X. F. Gong, R. Feng, and Z. Y. Zhu, "Ultrasonic investigation of the nonlinearity parameter B/A in biological media," J. Acoust. Soc. Am. **76**, 949–951 (1984).
[19] X. F. Gong, Z. M. Zhu, T. Shi, and J. H. Huang, "Determination of nonlinearity parameter for biological tissues using FAIS and ITD methods," J. Acoust. Soc. Am. **86**, 1–5 (1989).
[20] R. T. Beyer, "Nonlinear acoustics," Naval Ship Systems Command, New London, CT, 1974.
[21] Y. B. Yao, *Handbook of Physics Chemistry* (Publisher, Shanghai, China, 1985), pp. 593–606.
[22] J. Tong, Y. W. Dong, and H. Y. Zhao, "Study on the magnitude of the nonlineartiy parameter B/A inorganic liquids," Chin. Sci. Bull. **5**, 348–352 (1988).
[23] Z. M. Zhu, X. F. Gong, and J. H. Huang, "Measurements of the acoustic nonlinearity parameter B/A characterization using improved thermodynamic method," Acta Acust. (Beijing) **4**, 303–306 (1988).
[24] L. Bergmann, *Der Ultraschall und Scine Anwendung in Wissenschaft und Technik* (Verlag, Stuttgart, 1954).
[25] S. Prakash, S. K. Kor, and C. L. Singh, "Non-linearity acoustic parameter in higher alkanes," Acustica **27**, 28–31 (1972).
[26] Z. W. Qian, "Nonlinear acoustics in higher-order approximation," ACTA Physics Sinica, **4**, 670–675 (1995).

Reverberation of rapid and slow trills: Implications for signal adaptations to long-range communication

Marc Naguib[a)]
Department of Animal Behavior, University of Bielefeld, P.O. Box 100 131, 33501 Bielefeld, Germany

(Received 15 June 2002; revised 7 November 2002; accepted 25 November 2002)

Many acoustic signals in animals include trills, i.e., rapid repetitions of similar elements. Elements within these trills usually are frequency modulated and are degraded by reverberation during long-range transmission. Reverberation primarily affects consecutive elements with the same frequency characteristics and thus imposes a major constraint in the evolution of design and perception of long-range signals containing trills. Here transmission of frequency-unmodulated trills with different element repetition rates was studied. Trills were generated at different frequencies to assess frequency dependence of reverberation and then broadcast under three acoustic conditions—an open field and to assess seasonal changes in transmission properties, a deciduous forest before and after foliage had emerged. Reverberation was quantified at different positions within trills. The results show strong effects of vegetation density (season), transmission distance, frequency, element repetition rate, and element position within the trill on effects of reverberation. The experiments indicate that fast trills transmit less well than slow trills and thus are less effective in long-range communication. They show in particular that selection on trills should not act only on element repetition rate within trills but also on the trill duration as effects of reverberation increased with trill duration. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1539050]

PACS numbers: 43.80.Ev, 43.80.Lb, 43.80.Ka [WWA]

I. INTRODUCTION

Acoustic signals used in long distance communication are subject to considerable qualitative and quantitative changes during propagation through the environment (Piercy *et al.*, 1977; Wiley and Richards, 1978). Thus, the acoustic conditions of the transmission channel have profound implications for the evolution of signal design (Morton, 1975; Wiley, 1991; Tubaro *et al.*, 1993) and for the evolution on receivers perceptual mechanisms and response strategies (Wiley and Richards, 1978; Klump, 1996; Naguib and Wiley, 2001). Degradation of signals is habitat dependent and generally results in three structural changes in sound: reverberation, irregular amplitude fluctuations, and frequency-dependent attenuation, all of which differentially affect signal perception (Dooling, 1982; Naguib and Wiley, 2001). Although the general principles of sound degradation are well described (Wiley and Richards, 1978, 1982), there are still few studies that have separately quantified these aspects of signal degradation. Thus, still little is known about how sound degradation affects perception of particular signal structures, an issue important for understanding whether or not information coded in a signal is transmitted to a receiver under natural conditions.

In habitats with many sound-reflecting surfaces such as in forests, the most prominent form of degradation is reverberation which results in smearing of the temporal structure of the signal. Effects of reverberation are most prominent when subsequent elements are in the same frequency band, such as in trills, in particular when they consist of rapidly repeated elements. Trills are common in animal sounds and are particularly important in communication (Klump and Gerhardt, 1987; Stumpner and Ronacher, 1994; Podos, 1997). They vary considerably in element repetition rate and duration so that accumulating reverberation will differently affect perception of single elements depending on their rate and position within the trill. In addition, effects of reverberation on trills not only depend on the structure of the trill but also on vegetation density so that seasonal changes in vegetation will have profound effects on strength of reverberation. Such changes in transmission characteristics are important when adaptations of signals to the habitat are assessed. Nonmigratory songbirds, for instance, and migrants that arrive early at their breeding grounds in deciduous forests will encounter prominent changes in the acoustic condition of their habitats as foliage emerges. These changes in acoustic conditions are known to affect receivers' assessment of signals (Naguib, 1996) and adaptations of signals to the transmission characteristics also are likely to depend on the condition under which selection acts strongest on long-range communication.

Most studies that have quantified changes in the temporal structure of a signal used composite measures of degradation that are based on changes in the waveform of the signal (Gish and Morton, 1981; Dabelsteen *et al.*, 1993; Holland *et al.*, 1998; Brown and Handford, 2000) or that are based on spectrographic cross correlation (Fotheringham and Ratcliffe, 1995; Brown and Handford, 2000; Kime *et al.*, 2000). Although these measures on the waveform provide important holistic information on sound degradation, they are difficult to interpret from a perceptual perspective as they conflate the different forms of degradation. The waveform is affected by all degrading processes such as reverberation,

a)Electronic mail: marc.naguib@uni-bielefeld.de

frequency-dependent attenuation and irregular amplitude fluctuations.

Naguib and Wiley (2001) pointed out that ears are frequency analyzers and separately perceive changes in amplitude and frequency, so that separate measures of degradation in the different dimensions provide biologically more meaningful information in terms of how the degradation will affect signal perception and thus signal evolution. Animals' ears can have a high-frequency selectivity (Wilkinson and Howse, 1975; Dooling, 1982) so that reverberation at one frequency will not affect perception of a subsequent element at another frequency. Moreover, studies on auditory distance assessment have shown that the different components of degradation are processed separately in birds (Naguib, 1995, 1997a, b) and humans (Mershon and King, 1975; Little *et al.*, 1992). All of these approaches underline the separate biological importance of the different kinds of degradation.

So far, studies that directly quantified reverberation either used single tone pulses (Richards and Wiley, 1980; Waser and Brown, 1986; Naguib *et al.*, 2000) or single frequency-modulated notes from bird song and calculated the tail-to-signal ratio (Brown and Handford, 2000; Holland *et al.*, 2001) or the slope of the tail (Holland *et al.*, 2001), but effects of reverberation on subsequent elements to date have not be examined.

In this study I measured reverberation of trills with different repetition rates of elements and different frequencies broadcast over distances up to 120 m in three habitats differing seasonally in vegetation density. I focused on trills without frequency modulation in order to quantify perceptually salient parts of reverberation in those kinds of signals that are particularly affected by reverberation. Trills were expected to reverberate increasingly with distance and vegetation density. Reverberation was expected to increase particularly in dense vegetation with increasing element repetition rate and element position within a trill.

II. METHODS

A. Preparation of sound signals

All trills were synthesized on a PC using Cool Edit 2000 (Syntrillium Software Cooperation, Phoenix, USA) with a sampling rate of 44 100 Hz and 16-bit accuracy. In order to measure reverberation on the waveform of the signal without confounding effects of frequency modulation, trills were synthesized with frequency-unmodulated elements. The primary objective in measuring reverberation was to quantify effects of reverberation on closely spaced elements of similar frequency composition (trills). Synthesized trills were composed of eight elements of 10-ms duration and silent intervals between elements of either 10 ms (fast trills), 40 ms (medium trills), or 90 ms (slow trills). The element rates are common in song birds with the fast trills being significantly more frequent in open habitat species than in forest living species (unpublished data; also see Wiley, 1991). In order to assess effects of frequency on reverberation all trills were generated at 1, 3.5, and 7 kHz.

B. Transmission experiments

1. Habitat

Sound transmission experiments were conducted in November 2000 and in July 2001 at two transects at the same sites in a deciduous forests in Bielefeld, Germany. In November the experiments additionally were conducted on an open field but these experiments were not repeated in July as here no seasonal effects on reverberation were expected. In November vegetation was leafless and in July leaves had emerged fully. Acoustic conditions in terms of structure and density of sound-reflecting surfaces in November thus resembled those early in the breeding season when many birds start to sing but when leaves have not yet emerged. The acoustic conditions in July resembled those during the main breeding season. Trees in the forest mostly were beech and oak. Tree trunks varied in diameter with the diameter of the thickest trunks reaching 80 cm. Trees were spaced irregularly with about 3 to 5 m distance between closest trees. The ground mostly was plain and covered with leaves. There was little undergrowth and leaf carrying branches mostly grew above 5 m. The transmission experiments for the open habitat were conducted on a crop field with vegetation height of about 10 cm and with no further objects in the transmission path. The closest forest edge was over 50 m away from the microphone and loudspeaker positions. All experiments were conducted between 0600 and 0900 hours on days with calm weather.

2. Procedure

All signals were broadcast from a Toshiba Satellite 2210 CDT computer connected to a Blaupunkt MPA 2 amplifier and a Canton XS passive loudspeaker. The loudspeaker was connected to a telescopic pole and positioned at 4-m height which is a common perch height in singing birds. Sounds were broadcast with 90 dB, as measured at 1 m using a 2203 Brüel & Kjær precision sound level meter (C settings, fast response) on a 5-s 1-kHz tone of equal peak amplitude as all other signals in the broadcast sound file. Signals were rerecorded successively at 20, 40, 80, and 120 m using a Sennheiser ME 64/K6 microphone connected to a Sony TCD-D100 DAT recorder. The microphone is not fully omnidirectional but has similar sensitivity to sound arriving frontally and laterally and thus detected the presumably most important components of reverberation. The microphone was connected to a telescopic pole and, like the loudspeaker, was positioned at 4-m height to simulate the situation at which birds on their song posts perceive rivals' song. Trills were repeated twice at each distance in November and four times at each distance in July. In July the broadcast file was expanded to also include trills with different element repetition rates (40 and 90 ms) in order to assess how the expected stronger reverberation would affect trills with different element spacing. At all sites, sounds were broadcast over two transects. In the open field sounds were broadcast along the same transect but in opposite directions. In the forest sites, the two transects were perpendicular to each other with the loudspeaker being placed at the same position for both transects but faced in different directions (90°) for the two

transects. The same transects were chosen in July and November to permit analyses of changes in acoustic conditions with seasonally changing vegetation density.

C. Quantification of degradation

For analyses, sounds were resampled using Cool Edit on a PC with 44 100 Hz and 16 bit. All sounds were then bandpass filtered with filter settings specific to the frequency of the signal and normalized in amplitude. All bandpass filters were set at ±100 Hz from the carrier frequency of the signal (FFT size 8192, Hamming window, 60-dB attenuation). For measures of reverberation I measured the sound energy in trill elements and the energy in silent intervals between trill elements using Saslab Pro 3.95 software (Raimund Specht, Berlin). As measure of reverberation, the ratio of the energy in the 10-ms tone and the energy of the following 10 ms was calculated. The energy was calculated as the squared amplitude values (in Volts) multiplied by the sampling time (V*V*s). This *reverberation index* reflects the extent to which the silent intervals between successive notes is filled by energy resulting from reverberation by the previous note(s). The reverberation index was calculated for the first, fourth, and seventh elements in order to assess accumulating reverberation over the trill. An index of one indicates that the energy of the silent interval equaled that of the preceding tone, indicating that the silence between notes in the fast trills will be difficult to detect (Dooling, 1982; Okanoya and Dooling, 1990). The same procedure was used for measurements on trills with slower element repetition rates in order to permit a direct comparison among trills. As the silence between successive elements in trills with slower repetition rates (40 and 90 ms) was longer than the part measured (10 ms), an index of one in these trills does not indicate that the separate elements are difficult to resolve but indicates that the end of an element will be more difficult to determine. Using the 10-ms intervals for measurements in all trills better allowed assessment as to what extent reverberation accumulated from previous elements in the different trills.

D. Data analysis

All data were analyzed using multi-factorial ANOVAs. *Post hoc* pairwise comparisons where run using bonferioni *post hoc* tests, which adjusted *p*-values to multiple comparisons. In order to analyze seasonal effects of reverberation on the trills with element repetition rates of 10 ms, a four-factor ANOVA [2 (season)×4 (distance, 0, 20, 40, and 80 m)×3 (frequency, 1, 3.5, 7 kHz)×3 (element position within trill, first, fourth, seventh element) was used with four repetitions for July measures and two repetitions for November measures. The 120-m data could not be included in this comparison as the 3.5-kHz data in the experiments conducted in July were overlaid by songs from blue tits (*Parus caeruleus*) singing near the microphone positions so that sounds could not be analyzed accurately. Due to the additional trills with slower element repetition rates broadcast in July, a five factor ANOVA was used for a more detailed analysis of effects of reverberation on trills in July [4 (distance)×3 (element rate within trill, 10, 40, and 90 ms)×3 (frequency)×3 (element position within trill)] with four repetitions. For the same reason mentioned above, transmission distances of 120 m were not included in the analysis. Due to the problems encountered at 120 m in July, a third four-factor ANOVA was used on the data obtained in the forest in November for a more detailed analysis as here the 120-m data were included [2 (habitat)×5 (distance)×3 (frequency)×3 (element position within trill)] with two repetitions. Habitat effects in November were analyzed with a separate ANOVA as the data from the open field were not analyzed in more detail (due to the low level of reverberation). All statistic tests were performed with SPSS 10.01 on a PC. Signals were not included in the analysis when they were overlaid by excessive background noise. Only main and two-factor interactions are considered.

III. RESULTS

A. Reverberation of trills with 10-ms element repetition intervals

1. Comparison between the open habitat and the deciduous forest without foliage

Reverberation was significantly stronger in the deciduous forest without leaves (November) than on the open field ($F_{1,311}=61.315$, $P<0.001$). Reverberation was very low at the open field site and did not increase with distance whereas it increased significantly with distance in the forest [Figs. 1(a) and (b); also see analyses below].

2. Comparison between closed habitats with different vegetation densities

a. General. The overall analysis of variance over distances up to 80 m showed significant effects of season (deciduous forest with and without foliage) (Fig. 2) and transmission distance on reverberation of trills (Table I). In addition, the interactions between transmission distance and season, between season and frequency, and between distance and frequency, respectively, were significant (Table I). In other words, trills of different frequency were affected differently by reverberation over distance in the two habitats. Effects of reverberation showed a different pattern of increase with distance and with frequency before than after foliage had emerged, as detailed below.

b. Reverberation at different transmission distances. Effects of reverberation increased significantly with transmission distance before and after foliage had developed [Tables II and III; Figs. 1(b) and (c)]. However, the increase of reverberation with transmission distance in the forest differed among seasons, as indicated by the significant interaction between distance and season (Table I). In winter, reverberation did not increase significantly from 20 to 80 m but showed a strong increase at a transmission distance of 120 m [Fig. 1(b)]. In contrast, in summer, when foliage had developed fully, the increase of reverberation with transmission distance was more linear with significant increases over all transmission distances [Fig. 1(c)]. The highest level of reverberation reached in summer at transmission distances up to

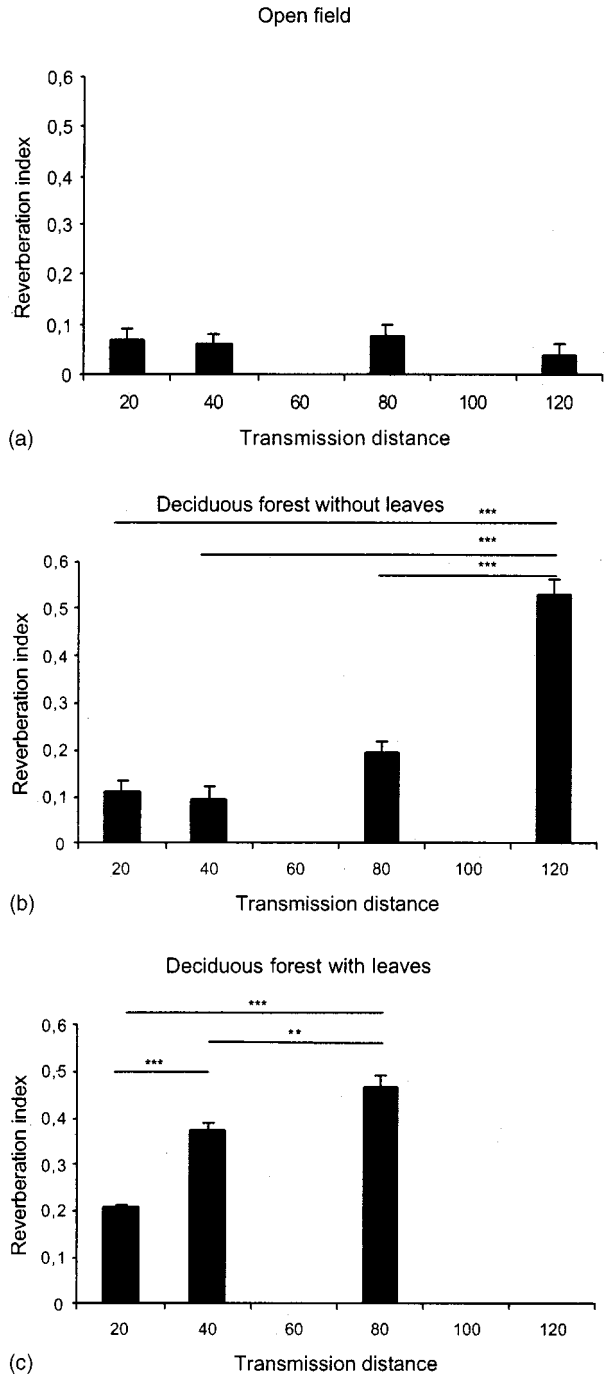


FIG. 1. Average reverberation indices for different transmission distances (a) on an open field in November, (b) in a deciduous forest in November, and (c) in a deciduous forest in July. Significant differences based on pairwise *post hoc* tests are indicated by stars (**$p<0.01$;***$p<0.001$).

80 m on average was about twice as high as that reached in winter, with indices at 80 m of 0.20 in winter and 0.47 in summer.

c. Reverberation at different element positions within a trill. Effects of reverberation increased significantly with element position within the trill in both seasons (Tables II and III). Reverberation increased significantly from the first to seventh element position before foliage had emerged (Fig. 3) and showed a significant increase already from the first to the fourth element after foliage had emerged ($p=0.02$) [Fig. 4(a)]. Moreover, the increase of reverberation with distance

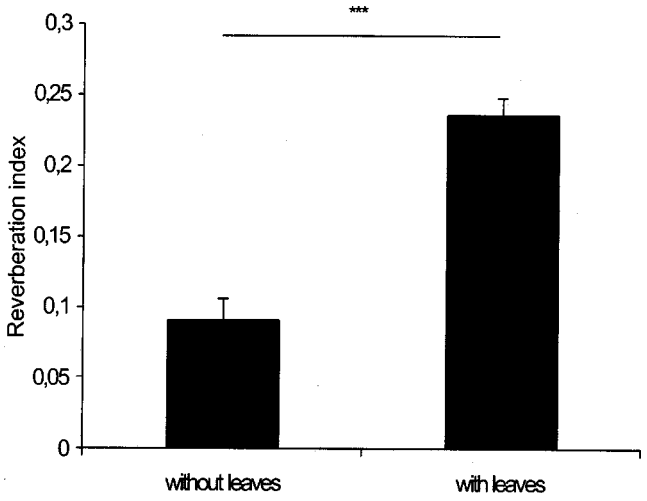


FIG. 2. Average reverberation index for the deciduous forest in winter before foliage had emerged and in summer with full vegetation. Significant differences based on pairwise *post hoc* tests are indicated by stars (***$p<0.001$).

was least for the first positions under both acoustic conditions, indicating that accumulating reverberation particularly affected subsequent element positions [Figs. 5(a) and (d)].

d. Reverberation at different frequencies. Overall, there was a significant effect of sound frequency on reverberation in the forest under both conditions (Tables II and III). In particular, there was a strong effect of season on reverberation at different frequencies [Figs. 6(a) and (b)]. In winter, reverberation decreased significantly with frequency, whereas in summer reverberation increased significantly with frequency. Reverberation at 1 kHz was similar under both conditions but higher frequencies reverberated stronger in summer with an average reverberation index at 7 kHz being almost three times as high than in winter [Figs. 6(a) and (b)].

B. Reverberation of trills with different element repetition rates

Trills with different element repetition rates accumulated reverberation to different extents within the first 10 ms following an element [Table III, Fig. 4(a)]. Trills with shorter element repetition rates showed significantly higher degrees of reverberation than did those with slower element repeti-

TABLE I. ANOVA on reverberation in 10-ms trills on combined data from the deciduous forest with and without foliage (November and July).

Source	d.f.	*F* ratio	*P*
Distance	3	34.026	<0.001
Season	1	58.395	<0.001
Frequency	2	0.750	0.473
Element-position	2	0.538	0.585
Distance×season	3	6.899	<0.001
Distance×frequency	6	3.440	0.003
Season×frequency	2	13.807	<0.001
Distance×element position	6	0.139	0.991
Season×element position	2	0.008	0.992
Frequency×element position	4	0.079	0.989

TABLE II. ANOVA on reverberation in 10-ms trills over transmission distances up to 120 m in the deciduous forest without foliage (November).

Source	d.f.	*F* ratio	*P*
Distance	4	37.707	<0.001
Frequency	2	8.044	<0.001
Element position	2	4.238	0.019
Transect	1	0.001	0.980
Distance×frequency	8	5.201	<0.001
Distance×element position	8	2.417	0.024
Distance×transect	4	4.546	0.003
Frequency×element position	4	0.643	0.634
Frequency×transect	2	7.810	0.001
Element position×transect	2	2.825	0.067

tion rates [Table III, Fig. 4(a)]. In trills with 10-ms element repetition intervals, reverberation increased with element position within a trill whereas it did not in trills with intervals of 40 and 90 ms, indicating that reverberation accumulated over several elements only in the fast trills [Fig. 4(b)]. This general pattern of reverberation also is reflected in the way it accumulated over distance. Only in the 10-ms trills reverberation increased more with increasing distance in the latter element positions than in the first element position. In trills with slower element repetition rates, the different element positions were not affected differently by reverberation at different distances [Figs. 5(b) and (c)].

IV. DISCUSSION

A. General

The experiments show that trills over distance reverberated differently depending on vegetation density, element repetition rate within the trill, frequency, and transmission transect within a given habitat. As predicted, they further show that silences between elements at later positions within trills with fast element repetition rates are most strongly affected by reverberation, supporting the prediction that not only element rate within trills but also that trill duration is

TABLE III. ANOVA on reverberation in trills with different element repetition intervals (10, 40, and 90 ms) over transmission distances up to 80 m in the deciduous forest with foliage (July).

Source	d.f.	*F* ratio	*P*
Distance	3	477.109	<0.001
Frequency	2	62.612	<0.001
Element position	2	5.970	0.003
Trill rate	2	7.069	0.001
Transect	1	18.036	<0.001
Distance×trill rate	4	10.711	<0.001
Distance×frequency	6	52.204	<0.001
Distance×element position	6	1.761	0.105
Distance×transect	3	17.444	<0.001
Element position×trill rate	4	5.271	<0.001
Element position×transect	2	5.819	0.003
Frequency×trill rate	4	4.995	0.001
Frequency×transect	2	3.73	0.025
Trill rate×transect	2	1.233	0.292
Element position×frequency	4	1.14	0.337

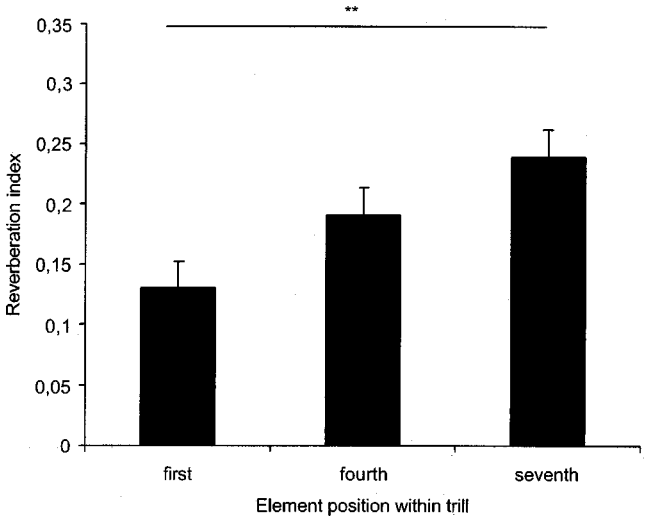


FIG. 3. Reverberation index for the different element positions within the trill in trills with element repetition rates of 10 ms broadcast in the deciduous forest before foliage emerged (November). Significant differences based on pairwise *post hoc* tests are indicated by stars (**$p<0.01$).

significantly affected by reverberation. The results further contribute to the general understanding of effects of reverberation on the evolution of signal structure and signal perception as effects of reverberation on trills had not been quantified yet in perceptually relevant frequency bands.

B. Effects of trill structure and transmission distance on reverberation in forests

The finding that trills with rapid element repetition rates reverberate more severely at later positions within the trill have important implications for understanding the evolution of trill structures in animal signals. Trills are common in orthopterans, anurans, and song birds and have been shown to be functionally important in female choice. Moreover, they are difficult to produce (Podos, 1997), and thus are likely to encode important information on a signaler's quality. As accumulating reverberation within the same frequency bands masks subsequent rapidly repeated elements with similar frequency, as shown here, such trills are not suited well for long-range communication in habitats with dense vegetation. In fact, previous studies on birds have shown that species in open habitats produce significantly faster trills than do species in forests (Wiley, 1991). Exceptions are Kentucky warblers (*Oporornis formosa*), a typical forest bird, which produces similar elements at short repetition intervals that are typical for open habitat species (Wiley and Godard, 1996), and blue tits which produce song types with fast trills regardless of vegetation density (Doutrelant *et al.*, 1999; Doutrelant and Lambrechts, 2001).

Wiley and Godard pointed out that trills could have both disadvantages and advantages for birds in forests—disadvantages because of difficulties in detecting trills or discriminating trill rates, advantages because of possibilities for enhancing ranging by receivers. Both processes should be affected by the length of the trill, because, as shown here, reverberation increases with position in a trill. We can thus predict that long trills would have nothing but disadvantages

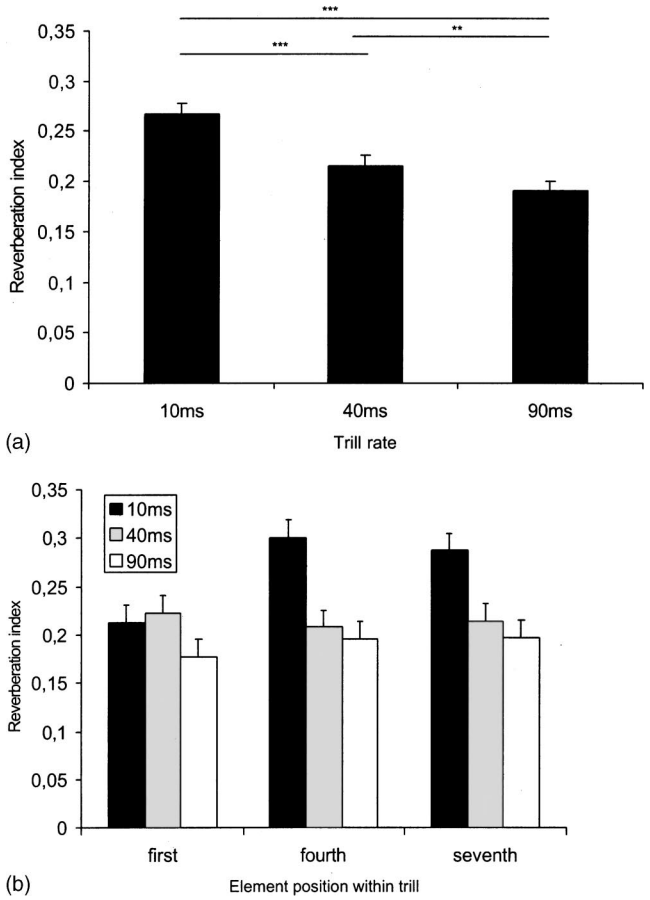


FIG. 4. Reverberation index for trills with different element repetition rates broadcast in a deciduous forest with foliage (a) overall means for the trills and (b) means separated by element positions within the trills with different element repetition rates. Significant differences indicated in (a) are indicated by stars ($^{**}p<0.01$; $^{***}p<0.001$).

for long-range communication in forests. Indeed few forest birds have long trills in their long-range songs (Wiley, 1991). On the other hand, predictions about the advantages and disadvantages of short trills are more difficult to make. It will thus be difficult to predict the optimal length for trills without further study. The accumulation of reverberation is probably not linear but instead levels off after a certain number of elements, which depends on element repetition rate. Furthermore, discrimination of trill rates and estimation of depth of amplitude modulation probably varies in complex ways with trill length. For ranging, a short trill might not accumulate enough reverberation or might not provide a long enough signal to estimate depth of amplitude modulation accurately. A long trill might accumulate too much reverberation and thus reduce the sensitivity of an estimate of amplitude modulation. Resolving these issues will require study of perception as well as transmission of trills.

C. Effects of habitat and seasonal changes of the acoustic environment on reverberation

Reverberation within fast trills accumulated with element position in the forest under both acoustic condition, i.e., without foliage and with foliage. However, the effect of element position became significant only after transmission distances of 120 m in the forest without foliage. When foliage had emerged in the same forest, the effect of reverberation on elements was stronger already at shorter distances. These seasonal changes of reverberation within a forest need to be considered when adaptations of signals to the acoustic conditions of a habitat and when responses of receivers to such degraded signals are studied. In particular the songs of year-round resident species have evolved under selection of vary-

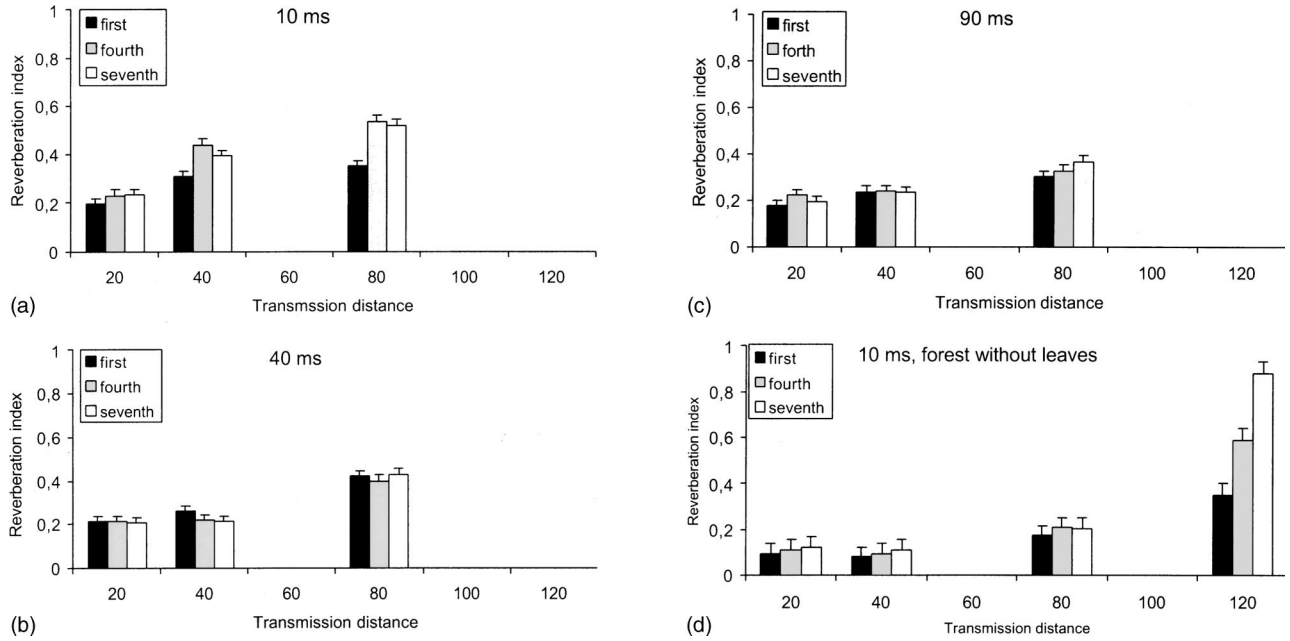


FIG. 5. Average reverberation indices at different distances for the different element positions within (a) trills with element repetition rate of 10 ms, (b) 40 ms, (c) 90 ms broadcast in deciduous forest with foliage and (d) trills with element repetition rate of 10 ms broadcast in deciduous forest before foliage emerged.

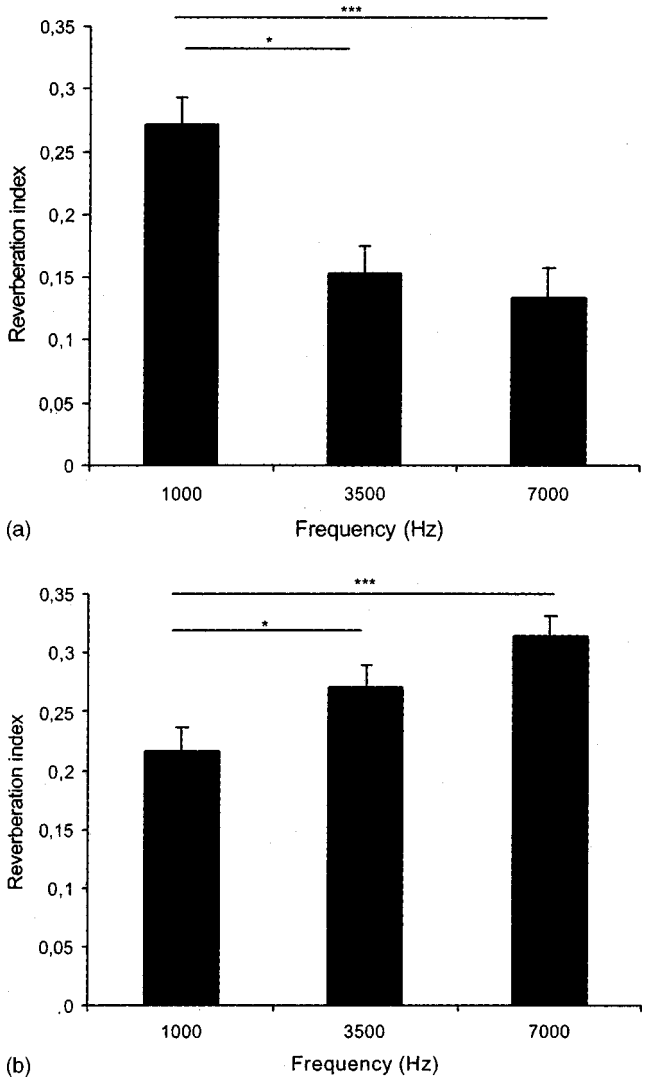


FIG. 6. Reverberation indices for trills with element repetition rates of 10 ms synthesized at different frequencies and broadcast in (a) deciduous forest before foliage emerged (November) and (b) deciduous with foliage (July). Significant differences are indicated by stars ($*p<0.01$; $**p<0.01$; $***p<0.001$).

ing acoustic conditions at different times of the year. As males in resident species usually start advertising territories before foliage has developed, these better acoustic conditions even may be more important in the evolution of signal design than the less favorite conditions later in the season when foliage has fully developed. There is evidence already that birds adapt response criteria to such seasonal changes in the acoustic conditions of their habitat. Carolina wrens showed differences in assessed auditory distance of songs before and after foliage had emerged even though songs were reverberated at the same level (Naguib, 1996). This experiment shows that such seasonal changes in reverberation not only are likely to influence signal structure but also the behavior of receivers. The finding that even the minor effects of reverberation on trills in the open habitat did not accumulate with distance or element position now provides a quantitative background for understanding the evolution of frequent use of fast long trills in song birds living in open habitats.

D. Effects of sound frequency on reverberation

In all habitats there was a strong interaction between sound frequency and transmission distance on extent of reverberation. Interestingly, these effects were different in the same forest with foliage than without foliage. Under both conditions reverberation was similar at 1 kHz, indicating that foliage in a forest with moderate vegetation density and with little undergrowth does not much affect reverberation at this frequency. Such an effect is expected as objects with sound reflecting surfaces best reflect frequencies with wavelengths smaller than the size of the object. A sound of 1 kHz has a wavelength of 33 cm and leaves have considerably smaller diameters of mostly less than 10 cm. In summer, foliage will considerably increase the surface that reflect sounds of higher frequencies, explaining the significantly stronger reverberation at higher frequencies compared to 1 kHz in the forest with fully developed foliage and compared to higher frequencies in the forest before foliage had emerged. Why higher frequencies before foliage emerged showed less reverberation than did 1 kHz is more difficult to interpret. Possibly reflection by the ground or additional defraction from tree trunks may have resulted in stronger reverberation at 1 kHz compared to higher frequencies.

E. Conclusions

The synthesized trills consisting of frequency-unmodulated elements allowed to separate effects of frequency-dependent attenuation from reverberation. Such a distinction is important when trying to understand how sound degradation affects signal perception and the evolution of signal design (Naguib and Wiley, 2001). Although the exact patterns of reverberation of trills will vary with habitat and the exact propagation path, the principles of how trills reverberate are likely to be a general characteristic of habitats with multiple sound reflecting surfaces. This study illustrates that rapid repetitions of elements with the same frequency suffer significantly more from reverberation than slower repetitions and thus will decrease the probability of signal detection and recognition. The data also show that it does not suffice to measure the spacing of elements within a trill in order to assess adaptations of signals to the acoustic conditions of a habitat. Future studies need to consider also the number of elements within a trills, as reverberation increases within trills consisting of rapidly repeated elements and thus will have different effects on trills with different duration.

ACKNOWLEDGMENTS

I thank Christine Brenninkmeyer, Benjamin Hundsdorfer, Nicole Meyer, and Annegret Walter for their help in conducting the transmission experiments and for part of the data analysis. I also thank Henrik Brumm, R. Haven Wiley, and two anonymous referees for comments on a previous version of the manuscript. The study was funded by a research innovation grant (FIF2) provided by the University Bielefeld.

Brown, T. J., and Handford, P. (**2000**). ''Sound design for vocalizations: quality in the woods, consistency in the fields,'' Condor **102**, 81–92.
Dabelsteen, T., Larsen, O. N., and Pedersen, S. B. (**1993**). ''Habitat-induced degradation of sound signals—Quantifying the effects of communication sounds and bird location on blur ratio, excess attenuation, and signal-to-noise ratio in blackbird song,'' J. Acoust. Soc. Am. **93**, 2206–2220.
Dooling, R. J. (**1982**). ''Auditory perception in birds,'' in *Acoustic Communication in Birds, Vol. 1* (Academic, New York), pp. 95–129.
Doutrelant, C., and Lambrechts, M. M. (**2001**). ''Macrogeographic variation in song—a test of competition and habitat effects in blue tits,'' Ethology **107**, 533–544.
Doutrelant, C., Leitao, A., Giorgi, H., and Lambrechts, M. M. (**1999**). ''Geographical variation in blue tit song, the result of an adjustment to vegetation type?'' Behaviour **136**, 481–493.
Fotheringham, R. J., and Ratcliffe, L. (**1995**). ''Song degradation and estimation of acoustic distance in black-capped chickadees (*Parus atricapillus*),'' Can. J. Zool. **73**, 858–868.
Gish, S. L., and Morton, E. S. (**1981**). ''Structural adaptations to local habitat acoustics in Carolina wren songs,'' Z. Tierpsychol. **56**, 74–84.
Holland, J., Dabelsteen, T., Paris, A. L., and Pedersen, S. B. (**2001**). ''Potential ranging cues contained within the energetic pauses of transmitted wren song,'' Bioacoustics **12**, 3–20.
Holland, J., Dabelsteen, T., Pedersen, S. B., and Larsen, O. N. (**1998**). ''Degradation of wren *Troglodytes troglodytes* song: Implications for information transfer and ranging,'' J. Acoust. Soc. Am. **103**, 2154–2166.
Kime, N. M., Turner, W. R., and Ryan, M. J. (**2000**). ''The transmission of advertisement calls in Central American frogs,'' Behav. Ecol. **11**, 71–83.
Klump, G. (**1996**). ''Bird communication in the noisy world,'' in Ecology and Evolution of Acoustic Communication in Birds, (Cornell U. P., Ithaca, NY), pp. 321–338.
Klump, G., and Gerhardt, H. C. (**1987**). ''Use of non-arbitrary acoustic criteria in mate choice by female gray tree frogs,'' Nature (London) **326**, 286–288.
Little, A. D., Mershon, D. H., and Cox, P. H. (**1992**). ''Spectral content as a cue to perceived auditory distance,'' Perception **21**, 405–416.
Mershon, D. H., and King, L. E. (**1975**). ''Intensity and reverberation as factors in the auditory perception of egocentric distance,'' Percept. Psychophys. **18**, 409–415.
Morton, E. S. (**1975**). ''Ecological sources of selection on avian sounds,'' Am. Nat. **109**, 17–34.
Nagulb, M. (**1995**). ''Auditory distance assessment of singing conspecifics in Carolina wrens: the role of reverberation and frequency-dependent attenuation,'' Anim. Behav. **50**, 1297–1307.
Naguib, M. (**1996**). ''Ranging by song in Carolina wrens *Thryothorus ludovicianus*: Effects of environmental acoustics and strength of song degradation,'' Behaviour **133**, 541–559.
Naguib, M. (**1997a**). ''Use of song amplitude for ranging in Carolina wrens, *Thryothorus ludovicianus*,'' Ethology **103**, 723–731.
Naguib, M. (**1997b**). ''Ranging of songs in Carolina wrens: effects of familiarity with the song type on use of different cues,'' Behav. Ecol. Sociobiol. **40**, 385–393; erratum **41**, 203.
Naguib, M., and Wiley, R. H. (**2001**). ''Estimating the distance to a source of sound: mechanisms and adaptations for long-range communication,'' Anim. Behav. **62**, 825–837.
Naguib, M., Klump, G. M., Hillmann, E., Griessmann, B., and Teige, T. (**2000**). ''Assessment of auditory distance in a territorial songbird: accurate feat or rule of thumb?'' Anim. Behav. **59**, 715–721.
Okanoya, K., and Dooling, R. J. (**1990**). ''Detection of gaps in noise by budgerigars (Melopsittacus undulatus) and zebra finches (Poephila guttata),'' Hear. Res. **50**, 185–192.
Piercy, J. E., Embelton, T. F. W., and Sutherland, L. C. (**1977**). ''Review of noise propagation in the atmosphere,'' J. Acoust. Soc. Am. **61**, 1403–1418.
Podos, J. (**1997**). ''A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae),'' Evolution (Lawrence, Kans.) **51**, 537–551.
Richards, D. G., and Wiley, R. H. (**1980**). ''Reverberations and amplitude fluctuations in the propagation of sound in a forest: implications for animal communications,'' Am. Nat. **115**, 381–399.
Stumpner, A., and Ronacher, B. (**1994**). ''Neurophysiological aspects of song pattern recognition and sound localization in grasshoppers,'' Am. Zool. **34**, 696–705.
Tubaro, P. L., Segura, E. T., and Handford, P. (**1993**). ''Geographic variation in the song of the rufous-collard sparrow in eastern Argentina,'' Condor **95**, 588–595.
Waser, P. M., and Brown, C. H. (**1986**). ''Habitat acoustics and primate communication,'' Am. J. Primatol **10**, 135–154.
Wiley, R. H. (**1991**). ''Associations of song properties with habitats for territorial oscine birds of eastern North America,'' Am. Nat. **138**, 973–993.
Wiley, R. H., and Richards, D. G. (**1978**). ''Physical constraints on acoustic communication in the atmosphere: implications for the evolution of animal vocalizations,'' Behav. Ecol. Sociobiol. **3**, 69–94.
Wiley, R. H., and Richards, D. G. (**1982**). ''Adaptations for acoustic communication in birds: sound transmission and signal detection,'' in *Acoustic Communication in Birds, Vol. 2* (Academic, New York), pp. 131–181.
Wiley, R. H., and Godard, R. (**1996**). ''Ranging of conspecific songs by Kentucky warblers and its implications for interactions of territorial males,'' Behaviour **133**, 81–102.
Wilkinson, R., and Howse, P. E. (**1975**). ''Time resolution of acoustic signals in birds,'' Nature (London) **258**, 320–321.