CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 115, No. 1 January 2004

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 115, NO. 1, JANUARY 2004

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 115, NO. 1, JANUARY 2004

CONTENTS—*Continued from preceding page*

ACOUSTICAL NEWS—USA

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

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

New Fellows of the Acoustical Society of America

Dezhang Chu—For contributions in acoustic scattering.

Timothy J. Foulkes—For the integration of acoustics in building design and to education in architectural acoustics.

David Feit honored by ASME International

David Feit, senior research scientist at the Carderock Division, Naval Surface Warfare Center (NSWCCD) in West Bethesda, has been awarded the Per Bruel Gold Medal for Noise Control and Acoustics by ASME International (The American Society of Mechanical Engineers). He is being recognized for advancing the understanding of structural-acoustic phenomena and subsequent noise and vibration control treatments related to submerged vehicles, high frequency noise radiation from fluid-loaded structures and the vibration of fuzzy structures.

Established in honor of Dr. Per Bruel, who pioneered the development of highly sophisticated noise and vibration measuring and processing equipment, the award recognizes eminent achievement and extraordinary merit in the field of noise control and acoustics. The medal was presented to Feit during the 2003 International Mechanical Engineering Congress and Exposition, which was held in Washington, DC, 16–21 November.

Feit has been employed at the NSWCCD since 1973. From 1973 to 1983, he supervised and managed a group of 40 engineers, scientists, and technicians who performed research and development on ships and submarines to achieve quieter vehicles. Since 1983 he has been the senior research scientist for structural acoustics with responsibility for conducting research into the complex problems related to the reduction of ship vulnerability to acoustic detection. This involves theoretical and applied research on the causes, mechanisms and means of reduction of submarine hull vibration and radiation, and echo reduction. During 1988–1990, he was on temporary assignment at the Office of Naval Research European Office in London, where he was the liaison scientist for acoustics and mechanics in Europe and the Middle East.

Prior to joining NSWCCD, he worked at Cambridge Acoustical Associates, Massachusetts (1964–1973). He has held adjunct professorial appointments at the Catholic University of America and George Washington University, Washington, DC; the U.S. Naval Academy (Annapolis, MD); and the University of Maryland, College Park.

During his career, Feit has authored/co-authored more than 50 articles in refereed journals. He co-authored the book *Sound, Structures and Their Interaction*, first published in 1972, which has since become internationally recognized as the classical reference on structural acoustics of submerged structures.

A Fellow of ASME, David Feit was a Rayleigh Lecturer in 1991 for the Noise Control and Acoustics Division. He is also a Fellow of the Acoustical Society of America (ASA) and a member of Sigma Xi. Among his other honors are ASA's Trent-Crede Medal (1999), NSWCCD's Sustained Superior Performance Award (1987) and David Taylor Award (1989), and the U.S. Navy's Meritorious Civilian Service Award (1982).

Feit received his A.B. at Columbia College in 1958. He earned his bachelor's, master's and Sc.D at the Columbia University School of Engineering in 1959, 1961, and 1964, respectively.

David Feit is the current Treasurer of the Acoustical Society.

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.

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: http://asa.aip.org].
12–14 July	Noise-Con 2004, Baltimore, MD [Institute of Noise

 0001-4966/2004/115(1)/1/2/$20.00

Control Engineering of the USA, Inc., INCE/USA Business Office, 212 Marston Hall, Iowa State Univ., Ames, IA 50011-2153; Tel.: 515-294-6142; Fax: 515-294-3528; E-mail: ibo@inceusa.org; WWW: http://www.inceusa.org/NoiseCon04call.pdf].

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: http://asa.aip.org].

Cumulative Indexes

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

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

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

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

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*
RICHARD S. MCGOWAN, *CReSS LLC, 1 Seaborn Place, Lexington, Massachusetts 02420*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
KEVIN P. SHEPHERD, *Mail Stop 463, NASA Langley Research Center, Hampton, Virginia 23681*
ROBERT C. WAAG, *Univ. of Rochester, Department of Electrical and Computer Engineering, Rochester, New York 14627*

6,596,206

43.28.Py GENERATION OF PHARMACEUTICAL AGENT PARTICLES USING FOCUSED ACOUSTIC ENERGY

David Soong-Hua Lee, assignor to Picoliter Incorporated
22 July 2003 (Class 264/9); filed 30 March 2001

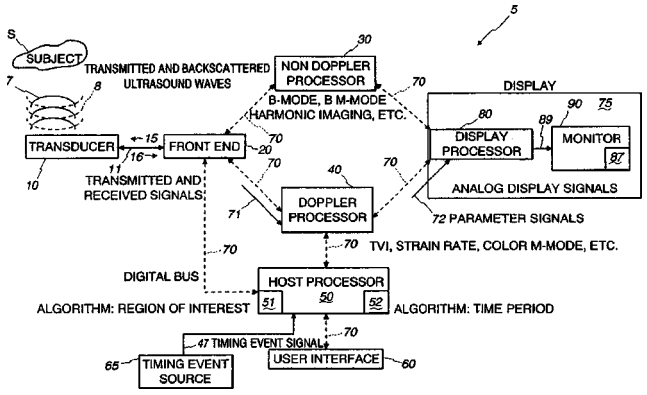
Focused acoustic energy is used to generate solid particles consisting of a pharmaceutical agent that can be dispensed in a solvent that may be an aqueous fluid, a nonaqueous fluid, or a supercritical fluid. Focused acoustic energy is used to eject a droplet of the solution, which is then directed into or through an antisolvent that upon admixture with the solution droplet causes the pharmaceutical agent in the droplet to precipitate.—DRR

6,592,522

43.35.Yb ULTRASOUND DISPLAY OF DISPLACEMENT

Steinar Bjaerum *et al.*, assignors to GE Medical Systems Global Technology Company, LLC
15 July 2003 (Class 600/443); filed 10 June 2002

This ultrasound system generates a color representation of a moving structure, such as cardiac wall tissue, within an area of interest and displays the results on a monitor. Each color characteristic corresponds to a set of signal values of the structure, such as velocity or rate of strain. A front end generates signals in response to backscattered ultrasound waves and a Doppler processor converts these to signals which represent the spatial patterns of the moving structure. A host processor incorporates a tracking function and a time integration function to generate tracked movement parameter profiles and displacement parameter values over a time period corresponding to sampled anatomical locations within the area of interest. These displacement parameter values are then mapped to color characteristic signals. A display processor maps the values of the color characteristic signals onto the displacement parameter values that represent the moving structure, thereby yielding a color representation on a video monitor.—DRR



6,599,245

43.35.Yb ULTRASOUND TRANSMISSION METHOD AND SYSTEM FOR SIMULATING A TRANSMIT APODIZATION

Qinglin Ma *et al.*, assignors to Siemens Medical Solutions USA, Incorporated
29 July 2003 (Class 600/437); filed 27 June 2000

This is a system for simulating a transmit apodization without the use of voltage regulation in line with each transducer element. In one embodiment, individual transducer elements are selectively turned on or off so that each transducer element either transmits a pulse or it does not. The selection of which elements transmit a pulse is determined in accordance with an apodization probability function (e.g., a Hamming function). Thus, fewer elements near the edges of the array transmit a pulse while a larger number of elements near the center do so. Consequently, the combined transmit pulses from all of the energized transducer elements simulate an apodized transmit beam. In another embodiment, each transducer element transmits a variable portion of a transmit pulse. Transducers at the outer edges of the array transmit a smaller part of a pulse than those elements located in the center of the array.—DRR

6,590,661

43.38.Ar OPTICAL METHODS FOR SELECTIVELY SENSING REMOTE VOCAL SOUND WAVES

J. Mitchell Shnier, Toronto, Ontario, Canada
8 July 2003 (Class 356/432); filed 19 January 2000

In the world of James Bond espionage, high-fidelity recordings can be made with a laser beam focused on a window 100 m distant. Sound waves vibrate the glass and these minute vibrations are superimposed on the reflected beam. This patent suggests that we don't even need the glass because laser interferometry or simple optical attenuation can detect disturbances in the air itself. By selecting wavelengths most affected by carbon dioxide or

water vapor such a system can be optimized for vocal pickup (people breathe). Although the patent appears to be purely speculative, it includes numerous calculations to support the basic theory.—GLA

6,584,206

43.38.Bs SPEAKER APPARATUS

Yoshio Ohashi, assignor to Sony Corporation
24 June 2003 (Class 381/191); filed in Japan 25 October 2000

The diaphragm of an electrostatic loudspeaker is divided into multiple individual electrodes. These are driven at different signal levels such that the driving forces are proportional to the surface displacement distribution of a single resonance mode.—GLA

6,590,992

43.38.Fx SPEAKER DEVICE

Takeshi Nakamura, assignor to Murata Manufacturing Company, Limited
8 July 2003 (Class 381/430); filed in Japan 24 March 1998

This is a strange patent. Some confusion probably results from the translation of technical terms from Japanese to English. Even so, the Background and Summary of the Invention sections are hard to follow. For example, high-quality conventional loudspeakers are designed to be as linear as possible. The patent describes the operation of such a speaker as ''...when a signal of large energy is input, the amplitude of the vibrator will increase, and as a consequence, sound pressure of the generated sound will increase.'' No problem, except that the next sentence reads, ''However, when sound pressure increases, the effect on the human body increases, which is unpleasant for the human listener.'' The remedy for this unpleasantness is a high-mass piezoelectric loudspeaker having its resonant frequency set higher than the upper limit of human hearing.—GLA

6,584,203

43.38.Hz SECOND-ORDER ADAPTIVE DIFFERENTIAL MICROPHONE ARRAY

Gary W. Elko and Heinz Teutsch, assignors to Agere Systems Incorporated
24 June 2003 (Class 381/92); filed 30 October 2001

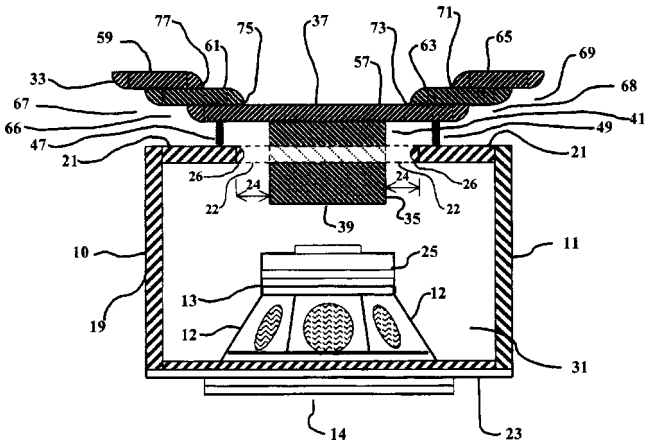
Signals from two back-to-back cardioid microphones, with variable gains plus fixed delays, can be combined to derive a variety of pickup patterns. The patent explains how this arrangement can be used as an adaptive array to track and null moving noise sources in both front and rear hemispheres.—GLA

6,578,660

43.38.Ja SPEAKER ENCLOSURE VENTURI EXPANDER

Ronald K. Taylor, Miami, Florida
17 June 2003 (Class 181/155); filed 18 June 2001

Among other loudspeaker novelties of the 1950s were enclosures with ''venturi'' vents. The name has resurfaced in this design which is intended to operate somewhat like a radial horn. Rearward sound waves from loudspeaker **13** are squirted through ring-shaped throat **22** and emerge from



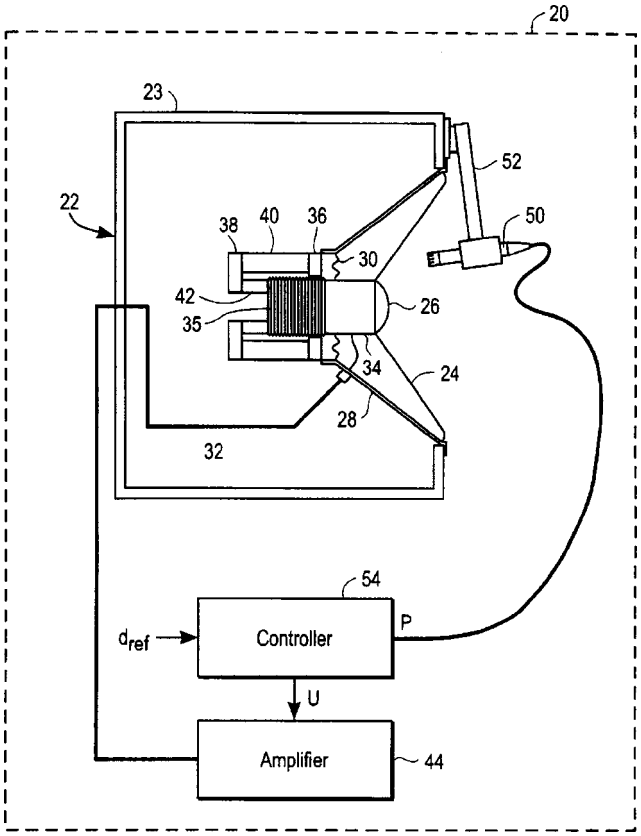
mouth **67** ''...so their reflections do not create cancellation by the mixing of out-of-phase rearward sound waves with the forward projected waves from the front of the speaker.'' No explanation is given for this miraculous behavior.—GLA

6,584,204

43.38.Ja LOUDSPEAKER SYSTEM WITH FEEDBACK CONTROL FOR IMPROVED BANDWIDTH AND DISTORTION REDUCTION

Khalid M. Al-Ali *et al.*, assignors to The Regents of the University of California
24 June 2003 (Class 381/96); filed 10 December 1998

Powered loudspeakers incorporating motional feedback have been available for at least 40 years. The feedback sensing element can be a pickup coil, a moving electrode, an accelerometer, or a microphone. This patent



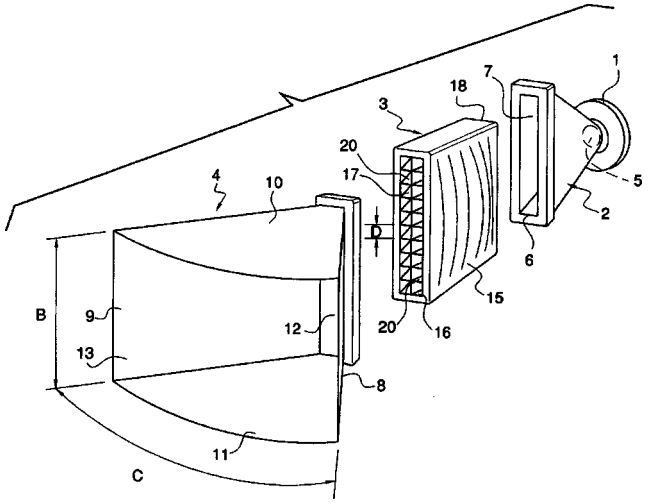
explains how a microphone and a few simple analog filters can be combined to form a stable feedback loop for a high-quality subwoofer. The text consists of a highly technical dissertation on applied servo control theory followed by 30 patent claims.—GLA

6,581,719

43.38.Ja WAVE SHAPING SOUND CHAMBER

Alan Brock Adamson, Scarborough, Ontario, Canada
24 June 2003 (Class 181/182); filed 27 July 2001

Line array fever grips the loudspeaker industry every 10 years or so. The current outbreak has prompted transducer designers to consider the practical problem of making a conventional high-frequency horn driver behave like a line source. One approach (described more than 50 years ago) is to fit the driver throat with a tightly packed cluster of equal-length flexible tubes. The opposite ends of the tubes are then arranged one above the other



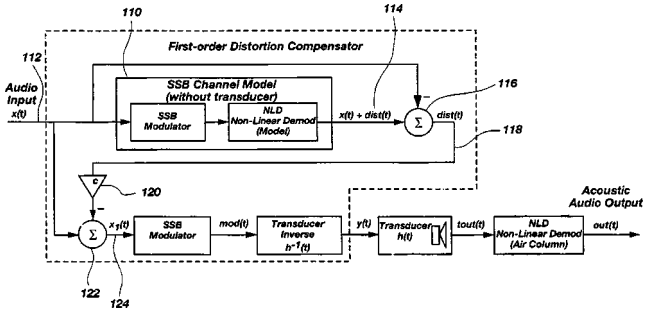
to form a line array. Several variants of this idea have recently been developed and patented. In this version, waveguide **2** is used to develop an arc-shaped wave front. The sound is then conducted through individually bowed channels of sound chamber **3** to drive conical horn **4**. The relative lengths of the channels can be set to achieve any desired degree of wave front flattening.—GLA

6,584,205

43.38.Ja MODULATOR PROCESSING FOR A PARAMETRIC SPEAKER SYSTEM

James J. Croft III *et al.*, assignors to American Technology Corporation
24 June 2003 (Class 381/98); filed 26 August 1999

A modulated ultrasonic signal can be used to create sound from thin air. The air itself acts as a demodulator. However, practical parametric loudspeaker systems are generally characterized by low efficiency and high dis-



tortion. This patent describes a signal processing system that includes dynamic predistortion and single sideband modulation to help alleviate both problems. Distortion canceling is applied to the audio signal alone to minimize bandwidth requirements.—GLA

6,590,993

43.38.Ja PANEL-SHAPED LOUDSPEAKER

Guido O. M. D'Hoogh and David M. E. Corynen, assignors to Koninklijke Philips Electronics N.V.
8 July 2003 (Class 381/431); filed in the European Patent Office 6 September 1999

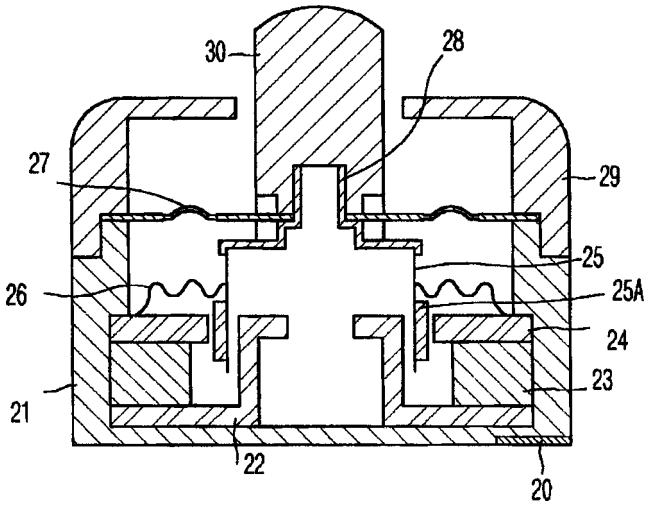
A corrugated cardboard panel is easier to bend parallel to the corrugations than at right angles to them. This patent asserts that a panel form loudspeaker can benefit from using a similar material. ''Surprisingly, it has been found that the loudspeaker according to the invention has a favorable acoustical behavior, both at low frequencies, central frequencies and high frequencies.'' More than half of the patent document is devoted to its 26 claims, most of which seem to restate the preceding claim in slightly different words.—GLA

6,483,917

43.38.Si TELEPHONE WITH A RECEIVER AROUSING BONE-CONDUCTION AND AIR-CONDUCTION HEARING

Kyeong Ok Kang *et al.*, assignors to Electronics and Telecommunications Research Institute
19 November 2002 (Class 379/433.02); filed in the Republic of Korea 19 August 1997

This cell phone is connected directly to a vibratorlike attachment which is said to stimulate the senses by bone conduction as well as by normal sound. A round bobbin is mounted directly onto a voice coil and is



supported much like a loudspeaker cone. The patent includes a brief description of the skull and which areas would be most useful as vibration-sensitive sites.—DLR

6,505,057

43.38.Si INTEGRATED VEHICLE VOICE ENHANCEMENT SYSTEM AND HANDS-FREE CELLULAR TELEPHONE SYSTEM

Brian M. Finn and Michael P. Nowak, assignors to Digisonix LLC
7 January 2003 (Class 455/569); filed 23 January 1998

A voice enhancement system for use in a vehicle is aimed at improving communication both within the vehicle and with hands-free operation of a cellular telephone. Multiple microphones are distributed throughout the vehicle cabin. A so-called microphone steering switch is provided which allows only one microphone to be switched ''on'' at any given time. However, the signal provided to the cell phone consists of this primary microphone

signal plus 20% of the signals from the microphones in the ''off'' state. This avoids the line appearing dead during gaps in speech. Noise reduction filters are used to improve the signal. Extensive details of circuit and filter designs are included.—KPS

6,580,794

43.38.Si ACOUSTIC ECHO CANCELER WITH A PEAK IMPULSE RESPONSE DETECTOR

Yoshihiro Ono, assignor to NEC Corporation
17 June 2003 (Class 379/406.08); filed in Japan 14 August 1998

This acoustic echo canceller is intended to cancel the acoustic coupling between a loudspeaker and microphone. An adaptive filter is used to estimate the echo signal so that it can be subtracted from the transmitted signal. To cancel residual echo after the subtraction, the magnitude of the tap signals of the filter is compared to the transmitted signal after subtraction. Further transmitted signal attenuation occurs if the tap signals are relatively large.—RSM

6,590,975

43.38.Si ECHO CANCELLER

Yoshihiro Ariyama, assignor to OKI Electric Industry Company, Limited
8 July 2003 (Class 379/406.09); filed in Japan 30 September 1998

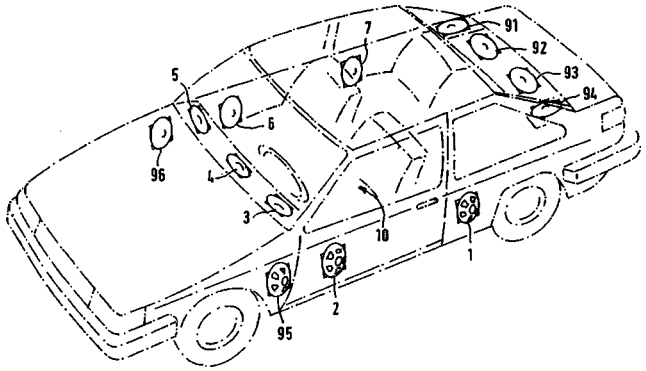
This echo canceller reduces the amount of calculation necessary to update adaptive filter tap coefficient values by reducing the number of tap coefficients. The eliminated tap coefficients correspond to the initial time delay in the echo path which is estimated by the impulse reponse.—RSM

6,584,202

43.38.Vk METHOD AND DEVICE FOR REPRODUCING A STEREOPHONIC AUDIOSIGNAL

Christoph Montag *et al.*, assignors to Robert Bosch GmbH
24 June 2003 (Class 381/63); filed in Germany 9 September 1997

A luxury sedan is fitted with at least 7, perhaps as many as 13, loudspeakers. Signals from left and right stereo channels are routed to each loudspeaker with predetermined amplitude and delay. Thus, the patent teaches, a coherent wave front can be generated that travels from the front



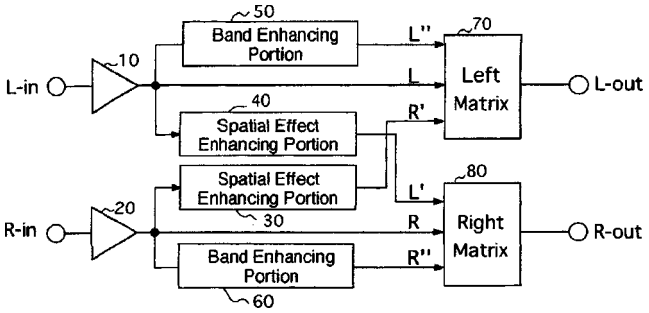
left to the rear right of the passenger compartment, for example. Additional delay taps can be used to control reflections. The theory is shaky, but 13 high-quality loudspeakers in a Mercedes should provide an impressive listening experience.—GLA

6,587,565

43.38.Vk SYSTEM FOR IMPROVING A SPATIAL EFFECT OF STEREO SOUND OR ENCODED SOUND

Pyung Choi, assignor to 3S-Tech Company, Limited
1 July 2003 (Class 381/98); filed in the Republic of Korea 13 March 1997

This patent describes yet another stereo enhancement method based on filtering and cross-coupling. In this case, signal processing is applied to left



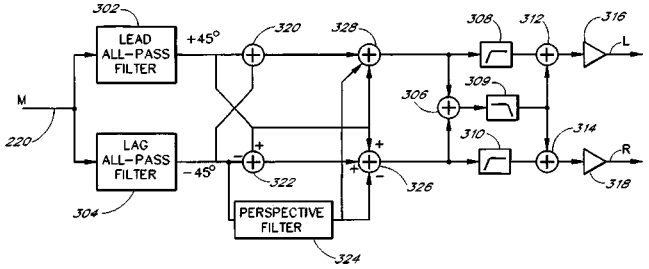
and right channels individually. The original left and right signals are ''enhanced'' at low frequencies but otherwise unmodified. The difference signal is ''enhanced'' in the high-frequency range.—GLA

6,590,983

43.38.Vk APPARATUS AND METHOD FOR SYNTHESIZING PSEUDO-STEREOPHONIC OUTPUTS FROM A MONOPHONIC INPUT

Alan D. Kraemer, assignor to SRS Labs, Incorporated
8 July 2003 (Class 381/17); filed 13 October 1998

Although the heyday of pseudo-stereo processors lies in the past, this patent argues that such devices can be used to advantage with electronic



musical instruments and concert sound systems. To minimize unwanted interaction with singers, the vocal range is accentuated in the L+R signal and correspondingly attenuated in the L−R signal.—GLA

6,602,229

43.40.Ng VIBRATING INJECTION NEEDLE

Ronald G. Coss, Newport Coast, California
5 August 2003 (Class 604/187); filed 24 August 2001

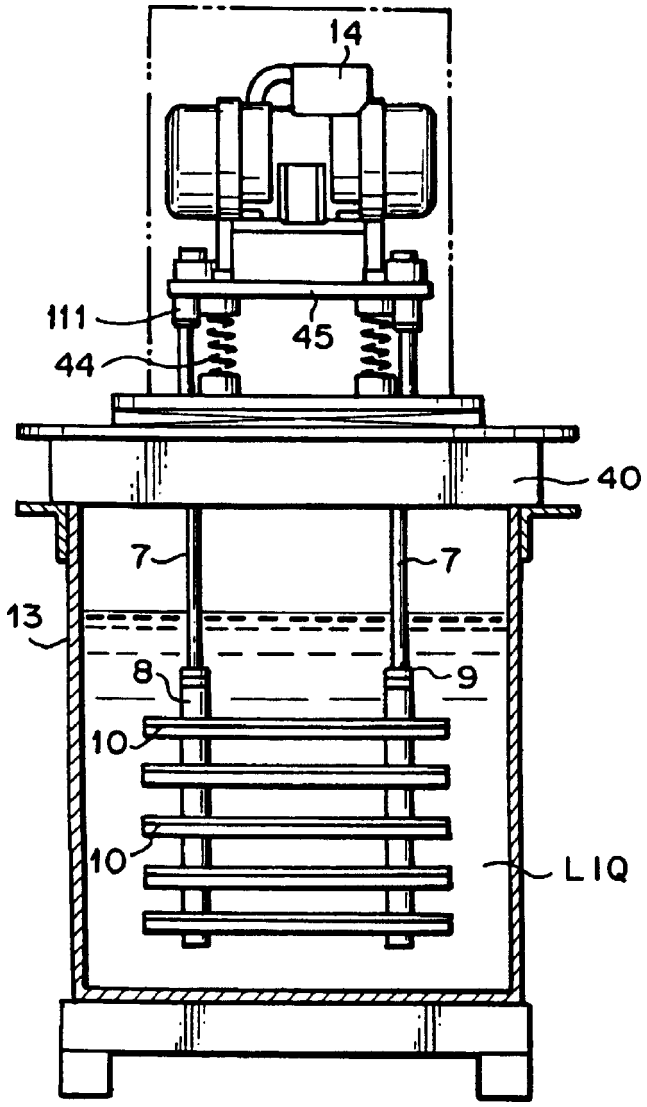
The purpose of vibrating an injection needle is to minimize the force of inserting medication into the tissue, thereby lessening the pain and discomfort in the patient. A portable battery-powered vibrating motor can be clipped onto any standard syringe to provide vibration to the needle. The clip attached to the syringe can be released from the driving motor so the syringe and the clip can be sterilized and the motor used on other syringes.—DRR

6,605,252

43.40.Rj VIBRATIONALLY STIRRING APPARATUS FOR STERILIZATION, STERILIZING APPARATUS AND STERILIZING METHOD

Ryushin Omasa, assignor to Japan Techno Company, Limited
12 August 2003 (Class 422/20); filed in Japan 2 May 2000

This vibrational liquid stirring apparatus contains a vibration motor **14**, a vibrating rod **7** connected to the motor, and vibrating vanes **10** affixed by members **9** to the vibrating rod. The vanes **10** have surfaces constructed of sterilizing metal and the fixing members **9** are fabricated from magnetic field generating materials. The vanes and fixing members are submerged in the



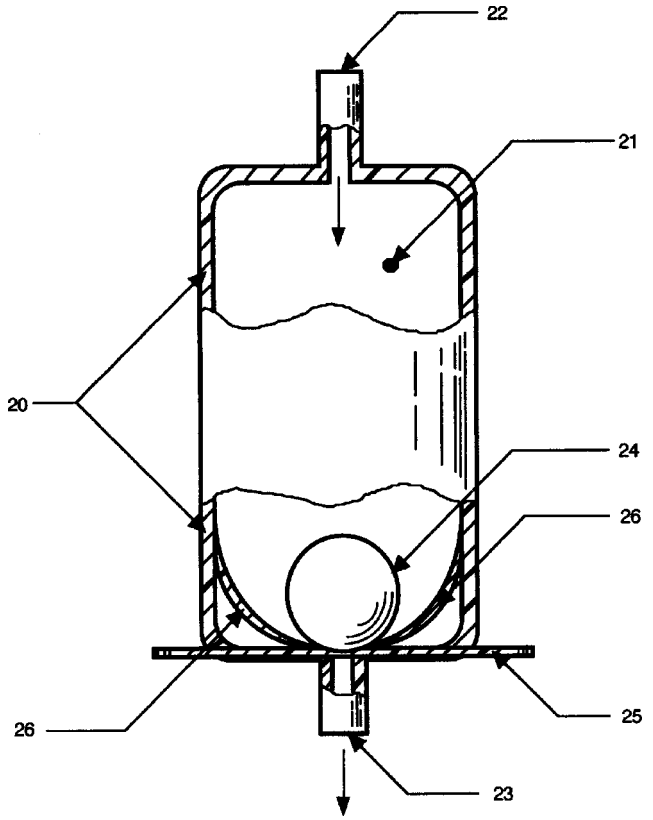
liquid **LIQ** in the treatment tank **13**. The vanes are vibrated with an amplitude of 0.1 to 15.0 mm at a frequency ranging from 200 to 1000 cycles per minute. The patent argues that when the fluid is vibrationally stirred, the liquid becomes sterilized through magnetic effects without the use of chemicals.—DRR

6,601,601

43.40.Vn ADJUSTABLE VIBRATION VALVE

Richard Tamian, Sparta, New Jersey
5 August 2003 (Class 137/38); filed 19 December 2001

This adjustable vibration valve is a device to regulate and control the flow of a fluid when the valve is vibrated by an outside source. Flow of the fluid is automatically stopped when the outside vibration does not occur. The device consists of a flow chamber with inlet and outlet ports, an adjustable valve seat between the lower interior volume of the flow chamber and the outlet port, and a specifically sized control ball that rests on the valve, thus



forming a seal when no vibration is present. When the device undergoes vibration from an outside source, the ball becomes dislodged from the valve seat, thus allowing the fluid present in the flow chamber to flow past the control ball, through the valve seat opening, and then out through the port.—DRR

6,598,479

43.40.Yq INTEGRAL MAGNET VIBRATION SENSOR

James C. Robinson *et al.*, assignors to CSI Technology, Incorporated
29 July 2003 (Class 73/658); filed 19 November 2001

This is a vibration sensor module that incorporates an integral magnet for nonpermanent attachment to a machine. In the preferred embodiment, the apparatus includes a housing containing a magnet, a vibration sensor, and a conditioning circuit. The vibration sensor detects machine vibration while the apparatus is magnetically attached to the machine. The conditioning unit optimizes the frequency response of the vibration sensor by reducing the gain of the sensor output at resonance.—DRR

6,598,480

43.40.Yq VIBRATION TESTING SYSTEM

Toshihiko Horiuchi *et al.*, assignors to Hitachi, Limited
29 July 2003 (Class 73/663); filed in Japan 17 September 2001

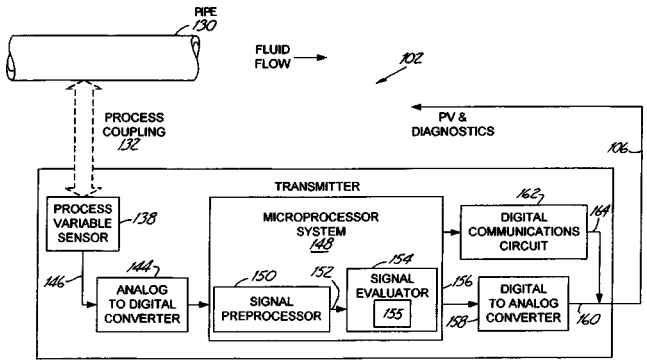
This patent describes a shaking testing system and test method that load deformation and/or force to a structure in order to evaluate the characteristics of the structure influencing its earthquake response, to verify the strength and reliability of the structure, or to obtain the vibration response of a massive structure. A main computer is provided for executing numerical simulation and a subcomputer controls each shaking machine. Data transfer between the main computer and subcomputer is performed by communications.—DRR

6,601,005

43.40.Yq PROCESS DEVICE DIAGNOSTICS USING PROCESS VARIABLE SENSOR SIGNAL

Evren Eryurek *et al.*, assignors to Rosemount Incorporated
29 July 2003 (Class 702/104); filed 25 June 1999

The patent relates to diagnostics of process devices, principally for use with industrial devices, whereby use is made of a process variable sensor signal. The methodology of process diagnosis involves isolating and evaluating vibration noise signals in the process fluid using a process variable



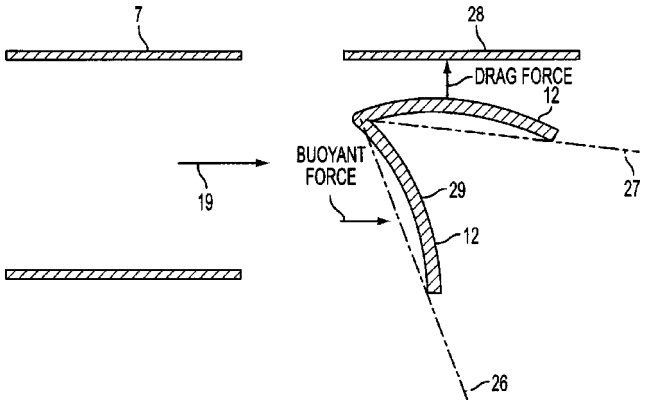
sensor signal. The signal preprocessor isolates components of the sensor signal relating to the operation of the process and produces an isolated signal output. A signal evaluator provides an output descriptive of the condition of the process control device as a function of the isolated signal.—DRR

6,499,562

43.50.Gf MUFFLER WITH VARIABLE SOUND-ABSORBING CHARACTERISTICS

Gerhard Elfinger *et al.*, assignors to Zeuna-Staerker GmbH & Company KG
31 December 2002 (Class 181/251); filed in Germany 29 July 1999

This is another automobile muffler designed with two flow paths, one for low engine speeds, the other for higher ones. The actuation method that switches between the two modes of operation relies on aerodynamic forces generated in the exhaust acting upon a spring-loaded device. Several



schemes are described, one of which (shown in the figure) utilizes the venturi effect to enhance stability. All of the designs are aimed at mechanical simplicity and avoid the use of sensors and actuators necessary in active systems.—KPS

6,520,285

43.50.Gf AUDIBLE TUNING APPARATUS FOR A MUFFLER

Mark Tobias, Franktown, Colorado
18 February 2003 (Class 181/241); filed 31 August 2001

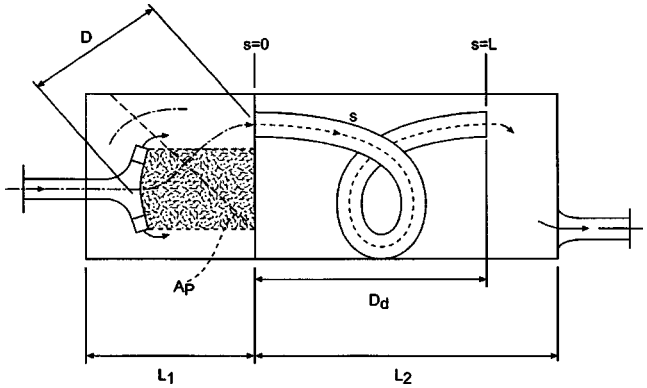
A muffler for a motorcycle is designed to have readily removable interior components. This is said to enable ''optimization'' of the engine's sound, including ''both increasing the volume level and causing the sound quality to emulate that of a hot rod.'' This capability is necessary because some motorcycle enthusiasts operate their vehicles at off-road events, not subject to noise regulations, but utilize public roadways to get to such events. Numerous removable muffler configurations are described.—KPS

6,520,286

43.50.Gf SILENCER AND A METHOD OF OPERATING A VEHICLE

Svend Frederiksen *et al.*, assignors to Silentor Holding A/S
18 February 2003 (Class 181/256); filed in Denmark 30 September 1996

Unlike most muffler patents that detail particular mechanical arrangements, this one describes a design method along with examples resulting from the application of the method. Both acoustic performance and pressure



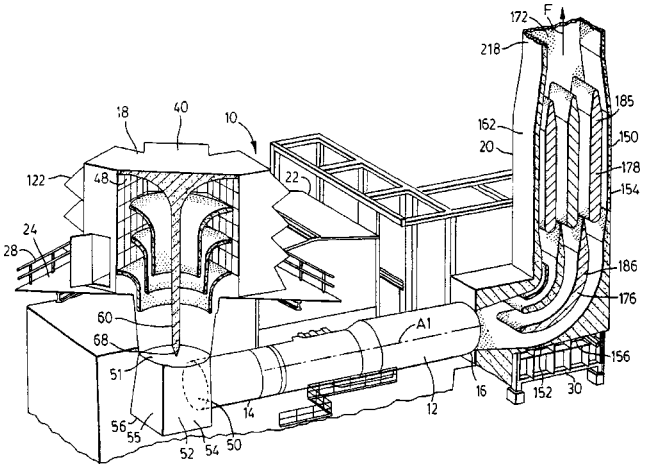
losses are included in the method. Furthermore, methods of operating the engine for noise reduction via control of fuel injection or air supply are also described. This extensive patent provides enormous detail to support its 285 claims!—KPS

6,537,490

43.50.Gf AIR INLET AND OUTLET SILENCER STRUCTURES FOR TURBINE

Ming Hui Han, assignor to M & I Heat Transfer Products Limited
25 March 2003 (Class 419/119); filed 30 May 2001

Methods for attenuating sound from stationary gas turbines, as typically used for power generation, are described for both inlet and exhaust noise. The inlet **40** contains a central, cylindrical structure **60** consisting of acoustical absorption material covered with perforated metal. In addition,



there are several annular airflow passages, the walls of which are also acoustically treated. At the exhaust of the turbine **16** is a large duct containing curved splitters. These too are acoustically treated and, because of the high-temperature exhaust, are composed of ceramic fiber or mineral wool covered with perforated stainless steel.—KPS

6,591,434

43.50.Gf SOUND DAMPENED SINK

Christopher M. De Cleene and Mark A. Romo, assignors to Kohler Company
15 July 2003 (Class 4/619); filed 4 February 2002

This reviewer finally got around to covering the proverbial kitchen sink. In the effort to lessen transmission of vibration along a countertop and noise annoyance to those in the nearby vicinity, pads, consisting of rubber-like material laminates, are affixed to the sink bottom and to a sidewall.—DRR

6,502,043

43.50.Ki PROCESS AND DEVICE FOR REDUCING THE SPECTRAL LINE NOISE INSIDE AN AIRCRAFT, ESPECIALLY A ROTATING-WING AIRCRAFT, IN PARTICULAR A HELICOPTER

Marc Dussac, assignor to Eurocopter
31 December 2002 (Class 702/56); filed in France 10 December 1999

Active control of noise within the cabin of a helicopter is achieved using either loudspeakers or force actuators. The control scheme is aimed at tonal noise for which a reference signal (e.g., shaft rotation) is readily available. Microphones provide the error signals and control is said to be achievable up to several kilohertz. This broadly written patent seems to encompass much of the active control technology developed for interior noise problems over the past two decades.—KPS

6,504,490

43.50.Lj VEHICLE DETECTION APPARATUS AND VEHICLE DETECTION METHOD

Koichiro Mizushima, assignor to Matsushita Electric Industrial Company, Limited
7 January 2003 (Class 340/943); filed in Japan 22 June 2000

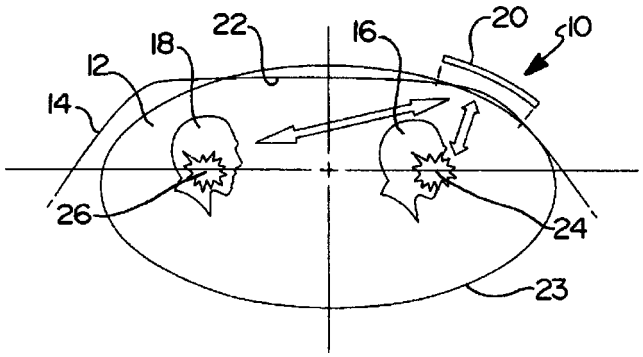
A method for detecting the passage of vehicles on a highway and determining their speed relies on an array of microphones of either one or two dimensions placed along the side of the road. Extensive practical details are provided such as desired microphone spacing, array sizes, and signal processing algorithms.—KPS

6,527,080

43.50.Lj PASSIVE SYSTEM FOR SPEECH ENHANCEMENT

Dafydd Geraint Davies *et al.*, assignors to Ford Global Technologies, Incorporated
4 March 2003 (Class 181/175); filed 2 December 2000

A passive system for speech enhancement in a vehicle consists of a curved, sound reflective element in or near the headliner. It is designed to enhance communication between passengers in the front and rear of the vehicle. The reflective surface **20** is a section of an ellipsoid **23** having its



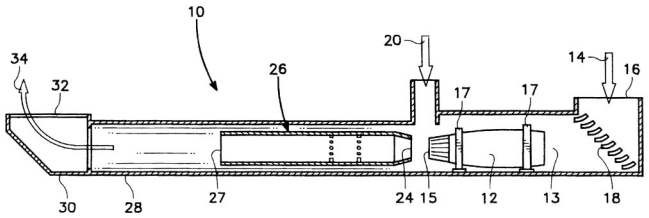
two foci at the passengers' head locations. The reflective surface is also curved in the dimension not shown in the figure. This arrangement is said to enhance speech levels by approximately 6 dB at frequencies above about 1 kHz.—KPS

6,497,137

43.50.Nm ANNULAR AFTER REACTOR WITH SOUND ATTENUATOR FOR USE IN A JET ENGINE TEST CELL AND TEST STAND

Norman Lewis Helgeson, assignor to The United States of America as represented by the Secretary of the Navy
24 December 2002 (Class 73/23.31); filed 7 May 2001

A design for a jet engine test cell is aimed at reducing both NOx emissions and noise. The exhaust **15** of jet engine **12** enters augmenter tube **28**. External air is drawn in through **20**, mixes with engine exhaust gases, and passes through diffuser **24**. This approach is designed to influence the



formation of the jet shear layer and reduce the jet velocity. Other noise control components of the design are acoustic treatment in the wall of tube

26 and placement of a honeycomb structure in the cross section downstream of the diffuser.—KPS

6,587,798

43.58.Dj METHOD AND SYSTEM FOR DETERMINING THE SPEED OF SOUND IN A FLUID WITHIN A CONDUIT

Alan D. Kersey *et al.*, assignors to Weatherford/Lamb, Incorporated
1 July 2003 (Class 702/50); filed 28 November 2001

This patent is short and clearly written. It is mainly concerned with oil and gas drilling. By measuring the speed of sound through a mixture of fluids—usually oil, gas, and water—the volumetric fraction of each component can be calculated. A method is described whereby established sonar techniques are used to solve what is essentially an inverse problem. The direction of fluid travel is already known. Moreover, no acoustic excitation is needed because sound-producing disturbances occur continuously as a consequence of fluid flow. By analyzing signals from two or more sensors and performing spatial/temporal decomposition, the slopes of spatial ridges can then be used to estimate phase velocity through the fluid.—GLA

6,496,734

43.66.Ts AUDITORY PROSTHESIS WITH AUTOMATED VOICE MUTING USING THE STAPEDIUS MUSCLE REFLEX

David K. Money, assignor to Cochlear Limited
17 December 2002 (Class 607/56); filed 24 April 2000

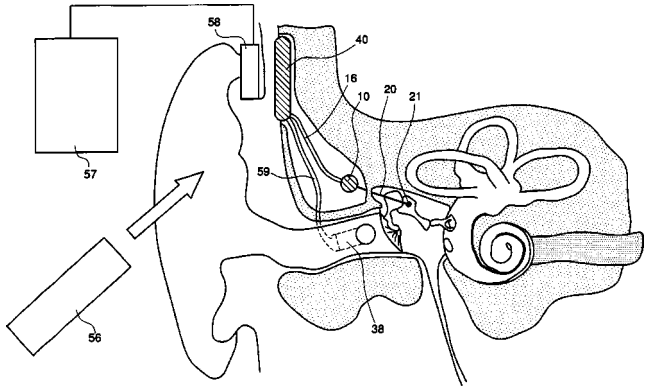
This auditory prosthesis includes a sensor attached to the stapedius muscle to detect activations of that muscle corresponding to loud ambient sounds as well as pending speech activity by the user. This muscle activity signal is then fed to the gain control circuitry of the device. The author argues that the stapedius reflex produces an ideal signal for use in gain control, primarily because of its relationship to speech activity as well as loud ambient sounds.—DLR

6,592,512

43.66.Ts AT LEAST PARTIALLY IMPLANTABLE SYSTEM FOR REHABILITATION OF A HEARING DISORDER

Friedemann Stöckert and Hans Leysieffer, assignors to Phonak AG
15 July 2003 (Class 600/25); filed in Germany 11 August 2000

This system contains at least one acoustic sensor for picking up acoustic signals and converting them into electrical audio sensor signals. An electronic unit combines audio signal processing and amplification. An elec-



tronic driver arrangement drives at least one electromechanical output transducer. The transducer contains an active electromechanical element and an output element for simulating an ossicle of a middle ear ossicular chain via a passive coupling element.—DRR

6,593,290

43.66.Ts TREATMENT OF INNER EAR HAIR CELLS

Wei-Qiang Gao, assignor to Genentech, Incorporated
15 July 2003 (Class 514/2); filed 4 November 1996

The purpose of this hair cell treatment is to induce, promote, or enhance the growth, proliferation, regeneration, or survival of inner ear tissue, particularly inner ear epithelial hair cells. Moreover, the application specifies methods, compositions, and devices for prophylactic and therapeutic treatment of inner ear disorders and conditions, especially impairments of hearing. The methods entail administration of insulinlike growth factor 1 (IGF-1) and/or fibroblast growth factor 2 (FGF-2) or their agonists.—DRR

6,594,346

43.66.Ts RELAY FOR PERSONAL INTERPRETER

Robert M. Engelke, assignor to Ultratec, Incorporated
15 July 2003 (Class 379/52); filed 14 February 2001

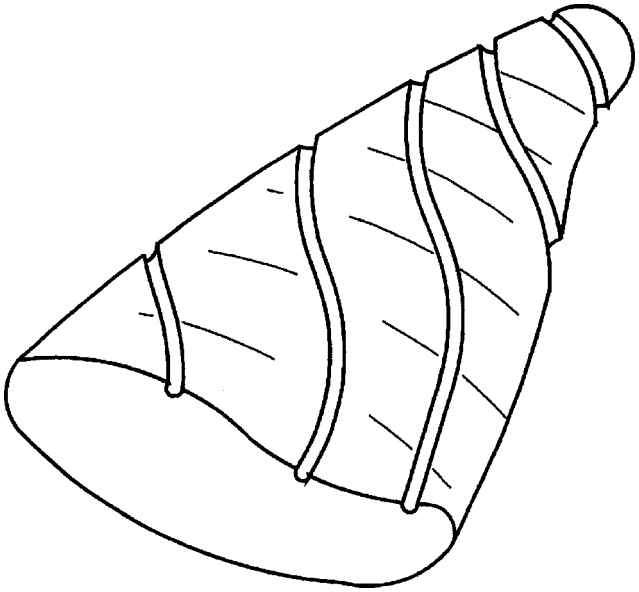
The intent of this relay is to facilitate communication through the telephone system between hearing users and users who need assistance in understanding voice communication. In order to overcome the speed limitation inherent in typing, the call assistant at the relay does not type most words, but revoices the words spoken by the hearing user into a computer using a voice recognition software package attuned to the voice of the call assistant. The text stream generated by the computer and the voice of the call assistant are both transmitted to the assisted user so that the latter person is supplied with a visual text stream to complement the voice communication. This system can be considered the strictly audio equivalent of the closed caption system used in television.—DRR

6,595,317

43.66.Ts CUSTOM-MOULDED EAR-PLUG DEVICE

Christoph Widmer *et al.*, assignors to Phonak AG
22 July 2003 (Class 181/135); filed 25 September 2000

In order to better absorb the dynamic movement of the ear canal that occurs during chewing and talking, earshells for custom-molded hearing aids are made very thin and reinforcing fins or ribs are added. Several



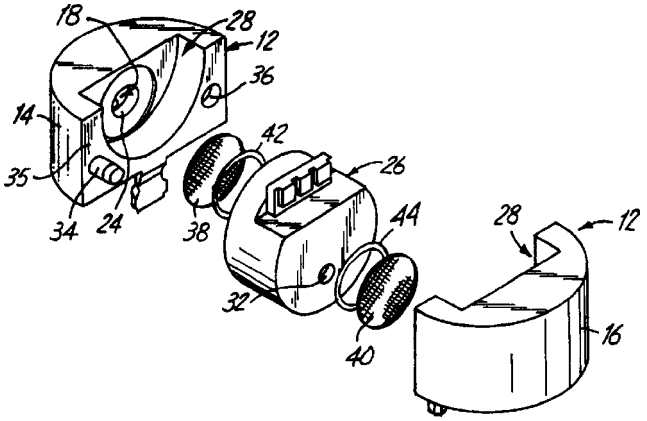
venting grooves may also be incorporated in the outside surface of the earshell to ensure aeration of the ear canal. Rapid prototyping via laser sintering or stereo lithography may be used to construct these detailed features.—DAP

6,597,793

43.66.Ts DIRECTIONAL/OMNI-DIRECTIONAL HEARING AID MICROPHONE AND HOUSING

Alex Darbut and Gerald D. Richels, assignors to Resistance Technology, Incorporated
22 July 2003 (Class 381/313); filed 6 August 1998

A manually operated mechanical switch enables an in-the-ear hearing aid wearer to toggle between directional and omnidirectional operating modes. The microphone is enclosed in a housing having two identical halves, each having an acoustic passage aligned with a microphone inlet.



The rotating switch opens and closes one of the inlet ports. When both ports are open to the two microphone inlets, the single directional microphone is in directional mode. When only one port is open to a microphone inlet, the microphone is in omnidirectional mode.—DAP

6,603,835

43.66.Ts SYSTEM FOR TEXT ASSISTED TELEPHONY

Robert M. Engelke and Kevin Colwell, assignors to Ultratec, Incorporated
5 August 2003 (Class 379/52); filed 23 August 2001

This is a bit of a rehash of the setup for providing captioned telephone service to hard-of-hearing persons as described in United States Patent 6,594,346, reviewed above.—DRR

6,603,858

43.66.Ts MULTI-STRATEGY ARRAY PROCESSOR

George Raicevich and Harvey Dillon, assignors to The University of Melbourne
5 August 2003 (Class 381/57); filed in Australia 2 June 1997

A signal processing system for hearing aids or cochlear implants determines the highest signal-to-noise ratio in the prevailing ambient noise via arrays of microphones. The processor connected to the first array produces a lower total noise output than the processor connected to the second array over a first range of noise floor values. The reverse is true over a second range of noise floor values. Gain of each processor output is varied in response to whether the noise floor signal is in the first or second noise floor range.—DAP

6,603,860

43.66.Ts APPARATUS AND METHOD FOR MONITORING MAGNETIC AUDIO SYSTEMS

Jon C. Taenzer and Jeffrey M. Sicurello, assignors to GN ReSound North America Corporation
5 August 2003 (Class 381/60); filed 29 July 1997

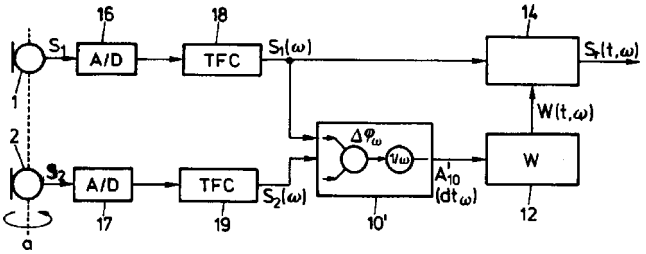
In order to test hearing devices that utilize an acoustic input signal and output a magnetic signal, a magnetic field detector is used to sense the level of magnetic output signal. The level of magnetic output signal is then converted to an acoustic signal. Thereafter, traditional hearing aid testing and listening checks may be performed.—DAP

6,603,861

43.66.Ts METHOD FOR ELECTRONICALLY BEAM FORMING ACOUSTICAL SIGNALS AND ACOUSTICAL SENSOR APPARATUS

Joseph Maisano and Werner Hottinger, assignors to Phonak AG
5 August 2003 (Class 381/92); filed in the European Patent Office 20 August 1997

A directional microphone system has directionality dependent on direction from which acoustical signals are received by at least two microphones, arranged at a predetermined mutual distance in the target direction.



A time delay detection unit has at least two inputs connected to the outputs of the microphones. The direction of the highest energy sound signals are assumed to be desired, while the signals arriving from other angles are assumed to be noise.—DAP

6,606,391

43.66.Ts FILTERBANK STRUCTURE AND METHOD FOR FILTERING AND SEPARATING AN INFORMATION SIGNAL INTO DIFFERENT BANDS, PARTICULARLY FOR AUDIO SIGNALS IN HEARING AIDS

Robert Brennan and Anthony Todd Schneider, assignors to dspfactory Limited
12 August 2003 (Class 381/316); filed 2 May 2001

A generalized method of digital filtering for hearing aid amplification is described. The audio signal is filtered into several uniformly spaced analysis frequency bands. After processing, a synthesis filterbank provides means for recombining the separately processed signals. The number of frequency bands and the bandwidth of the analysis and synthesis filterbanks are selectable. Coefficients for and the relationship between analysis and synthesis windows may also be chosen.—DAP

6,493,576

43.66.Yw METHOD AND APPARATUS FOR MEASURING STIMULUS-EVOKED POTENTIALS OF THE BRAIN

Franz Dankwart-Eder, assignor to Erich Jaeger GmbH
10 December 2002 (Class 600/544); filed in Germany 17 June 1996

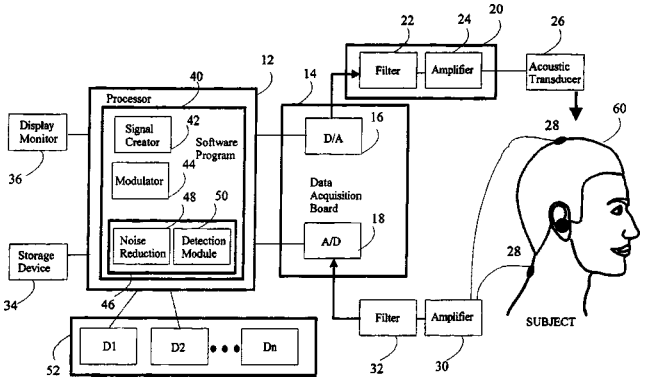
During surgery, it is necessary to quickly and accurately determine any decrease in the depth of anesthesia so that measures can be taken to prevent the patient from becoming conscious. For this purpose, clicks are applied via headphones and brain stem acoustically evoked potentials (BAEPs) and midlatency auditory evoked potentials (MLAEPs) are measured and analyzed. It is known as prior art that the amplitude of MLAEP-specific peaks decreases with increasing levels of anesthesia, while their latencies expand. Much of the patent text deals with details of the circuitry required to distinguish between BAEP and MLAEP signals.—DLR

6,602,202

43.66.Yw SYSTEM AND METHODS FOR OBJECTIVE EVALUATION OF HEARING USING AUDITORY STEADY-STATE RESPONSES

Michael Sasha John and Terence W. Picton, assignors to Baycrest Centre for Geriatric Care
5 August 2003 (Class 600/559); filed 18 May 2001

This apparatus, intended to evaluate a subject's hearing by recording steady-state auditory evoked responses, generates a steady-state auditory evoked potential stimulus and presents the stimulus to the subject. Evoked potentials are sensed simultaneously with stimulus presentation and the system determines whether the sensed potentials contain responses to the stimulus. The stimulus may include an optimum-vector-combined amplitude



and frequency modulation signal adjusted to evoke responses with increased amplitudes, an independent amplitude and frequency modulation signal, and a signal whose envelope is modulated by an exponential modulation signal. The apparatus may also be adapted to conduct phase-weighted T-tests or phase zone techniques, latency tests, AM/FM discrimination tests, rate sensitivity tests, aided hearing tests, etc.—DRR

6,487,533

43.72.Ja UNIFIED MESSAGING SYSTEM WITH AUTOMATIC LANGUAGE IDENTIFICATION FOR TEXT-TO-SPEECH CONVERSION

Henry C. A. Hyde-Thomson and Roger Liron, assignors to Avaya Technology Corporation
26 November 2002 (Class 704/260); filed 10 January 2000

This system is described as a voice gateway server which provides voice message services to a community of subscribers. When a text message arrives for a subscriber, the system performs a trigram analysis and compares the resulting trigram statistics to sets of trigram frequencies for several different languages. Based on this result, one of several language-specific speech synthesis algorithms is activated to generate output in the appropriate language.—DLR

6,490,563

43.72.Ja PROOFREADING WITH TEXT TO SPEECH FEEDBACK

Hsiao-Wuen Hon *et al.*, assignors to Microsoft Corporation
3 December 2002 (Class 704/260); filed 17 August 1998

This patent describes a proofreading system primarily designed to handle the multiple written forms commonly encountered in eastern Asian languages. Facilities are provided for conversions between, and speech synthesis from, several styles of character sets in both Chinese and Japanese, as well as several alphabetic spelling systems in each of the languages. A speech recognition module also allows the proofreader's speech to be inserted into the document in the appropriate written form. The patent does not discuss whether Korean texts could also be handled with similar techniques.—DLR

6,480,819

43.72.Ne AUTOMATIC SEARCH OF AUDIO CHANNELS BY MATCHING VIEWER-SPOKEN WORDS AGAINST CLOSED-CAPTION/ AUDIO CONTENT FOR INTERACTIVE TELEVISION

Robert Boman and Jean-Claude Junqua, assignors to Matsushita Electric Industrial Company, Limited
12 November 2002 (Class 704/9); filed 25 February 1999

This speech recognizer is set up to process target words from the user which will be used to search one or more incoming television channels. A parser and semantic analyzer determine whether the recognized speech is a content search target or a system command. Incoming audio channels may also be directed to the speech recognizer, allowing words to be found in the audio stream. There is no mention of an image processor to decode text from the video image. Upon detecting a match, the incoming television signal may be displayed, recorded, or both. The parser is said to be able to handle input such as ''I want to record the Seinfeld re-run tomorrow night.''—DLR

6,484,136

43.72.Ne LANGUAGE MODEL ADAPTATION VIA NETWORK OF SIMILAR USERS

Dimitri Kanevsky *et al.*, assignors to International Business Machines Corporation
19 November 2002 (Class 704/9); filed 21 October 1999

This network-based language recognition system accepts inputs consisting of speech, handwriting, gestures, or ''other similar actions'' from multiple users. Inputs are clustered according to similarities among various features. Both user model updates and user activity patterns are distributed over the network to all other user models in the same cluster. As the system learns about specific contextual domains, this information is also shared among other users who deal with similar input issues.—DLR

6,487,530

43.72.Ne METHOD FOR RECOGNIZING NON-STANDARD AND STANDARD SPEECH BY SPEAKER INDEPENDENT AND SPEAKER DEPENDENT WORD MODELS

Lin Lin and Ping Lin, assignors to Nortel Networks Limited
26 November 2002 (Class 704/244); filed 30 March 1999

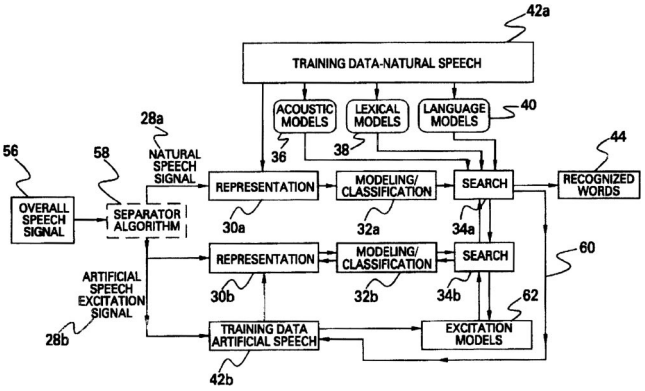
This speaker-independent speech recognizer attains that status by using a combination of speaker-dependent and speaker-independent word representations. Word features are stored in one or the other model according to the patterns of feature matching characteristics across speakers. A word can then be recognized if it is found in either the local dependent model or in the global independent model.—DLR

6,487,531

43.72.Ne SIGNAL INJECTION COUPLING INTO THE HUMAN VOCAL TRACT FOR ROBUST AUDIBLE AND INAUDIBLE VOICE RECOGNITION

Carol A. Tosaya and John W. Sliwa, Jr., both of Los Altos, California
26 November 2002 (Class 704/246); filed 6 July 1999

This is a novel technique for improving speech recognition, applicable in certain, specific circumstances. An artificial vocal tract excitation signal containing a known spectral content is introduced into the vocal tract of the speaker. During recognition processing, the known spectral content can be



detected and used to provide a more precise analysis of the speaker's articulations. An interesting aspect of the system is that the artificial excitation signal need not be audible, but could lie outside of the hearing frequency range. This allows a variety of interesting applications, such as silent speech, detectable only by the nearby recognizer.—DLR

6,487,532

43.72.Ne APPARATUS AND METHOD FOR DISTINGUISHING SIMILAR-SOUNDING UTTERANCES SPEECH RECOGNITION

Koen Schoofs and Guido Gallopyn, assignors to ScanSoft, Incorporated
26 November 2002 (Class 704/251); filed 24 September 1998

This speech recognizer includes a method of resolving ambiguities among sets of similar-sounding words. When an unresolved recognition situation arises, the user is asked to provide one or more hints, which may take the form of a context reference, a partial spelling, another related word, or any of a variety of ''linguistic properties.'' Most of the patent text appears to deal with general problems of using a recognizer to edit text materials.—DLR

6,499,011

43.72.Ne METHOD OF ADAPTING LINGUISTIC SPEECH MODELS

Bernd Souvignier and Andreas Kellner, assignors to Koninklijke Philips Electronics N.V.
24 December 2002 (Class 704/240); filed in Germany 15 September 1998

In cases where phonetic feature analysis results are not decisive in determining the correct word corresponding to an acoustic signal, an *N*-best list of possible alternatives is constructed. Each item in the list is assigned a probability value based on the phonetic similarity and a weight which represents its frequency of occurrence in the application. These measures are considered during subsequent syntactic and semantic analysis steps to determine the most suitable item from the *N*-best list.—DLR

6,592,521

43.80.Qf MEDICAL DIAGNOSTIC ULTRASOUND SYSTEM AND METHOD FOR POWER CONTROL

Joseph A. Urbano *et al.*, assignors to Acuson Corporation
15 July 2003 (Class 600/441); filed 1 March 2000

The purpose of this system is to provide consistent power while reducing power consumption by an ultrasound system. The ultrasound system is provided with an uninterruptible power supply. When the power from an outlet falls below a threshold value, is of insufficient quality, or is otherwise unavailable, the uninterruptible power supply comes into play to supply the requisite power. For mobility, the uninterruptible power supply precludes the necessity for a power outlet to energize an ultrasound system. In order to lessen the drain on the uninterruptible power supply or other source of power, a power reduction system is incorporated into the ultrasound system.—DRR

6,592,524

43.80.Qf TRANSMIT BEAMFORMER DELAY ARCHITECTURE AND METHOD FOR DIAGNOSTIC MEDICAL ULTRASOUND

Joseph A. Urbano *et al.*, assignors to Siemens Medical Solutions USA, Incorporated
15 July 2003 (Class 600/447); filed 22 December 2000

This involves a transmit beamformer for a medical diagnostic ultrasound system. A specific delay is implemented for each beamformer channel. The transmit beamformers generate electrical waveforms in a plurality of channels for relay to associated elements of a transducer array. By applying different delays to different channels, a beam of ultrasound energy is steered and focused.—DRR

6,595,316

43.80.Qf TENSION-ADJUSTABLE MECHANISM FOR STETHOSCOPE EARPIECES

George Cybulski *et al.*, assignors to Andromed, Incorporated
22 July 2003 (Class 181/131); filed 18 July 2001

This is a spring-actuated adjustable mechanism for varying the tension of the earpieces of a stethoscope against the ears.—DRR

6,595,923

43.80.Qf ULTRASONIC DOPPLER FLOW PHANTOM

David M. Sjoblom, assignor to Koninklijke Philips Electronics N.V.
22 July 2003 (Class 600/437); filed 14 November 2001

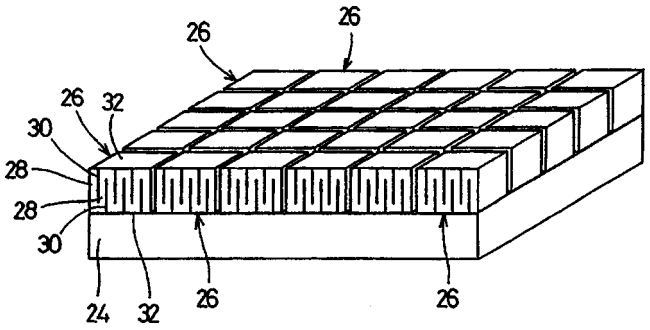
Phantoms are devices to evaluate and calibrate the performance of ultrasonic systems. A phantom is a boxlike device containing tissue-mimicking material possessing an acoustic response matching that of the human body. In the embodiment described in this patent, an ultrasonic Doppler flow phantom is formed of tissue-mimicking material containing a plurality of fluid flow paths through which fluid is pumped. The fluid flow paths are constructed of tubing and each path extends over a different portion of the depth of the phantom and at a different angle. Each path is individually accessible by a probe located on the simulated skin surface at the top of the phantom, thus allowing the performance of the probe to be evaluated over a wide range of depths from shallow to rather deep locations and over several flow angles.—DRR

6,603,240

43.80.Qf SENSOR ARRAY, METHOD FOR MANUFACTURING SENSOR ARRAY, AND ULTRASONIC DIAGNOSTIC APPARATUS USING THE SAME

Yoshiaki Kohno and Masato Yabuuchi, assignors to Murata Manufacturing Company, Limited
5 August 2003 (Class 310/334); filed in Japan 27 September 1999

This sensor array includes a substrate and multiple piezoelectric oscillators affixed to a main surface of the substrate in a matrix format. Each of the oscillators includes a number of piezoelectric layers laminated in a direction parallel to the substrate's main surface, inner electrodes deployed between the piezoelectric layers, and outer electrodes formed on the end faces of the piezoelectric layers. The procedures for manufacturing this sensor array are also covered in the patent.—DRR



6,605,043

43.80.Qf DIAGNOSTIC MEDICAL ULTRASOUND SYSTEMS AND TRANSDUCERS UTILIZING MICRO-MECHANICAL COMPONENTS

William R. Dreschel *et al.*, assignors to Acuson Corporation
12 August 2003 (Class 600/459); filed 30 December 1998

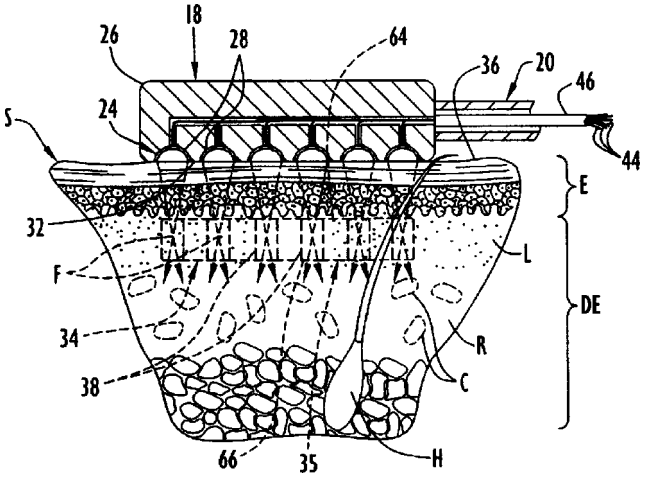
This patent was granted for an application that appears to be rather generic and not necessarily original in claiming the use of any micro-mechanical component in an ultrasound system. More specifically, this refers to the use of micro-relays, micro-switches, or inductors in the transducer probe head or in the transducer connector, as well as in other system transducer connections, or anywhere else in the system.—DRR

6,595,934

43.80.Sh METHODS OF SKIN REJUVENATION USING HIGH INTENSITY FOCUSED ULTRASOUND TO FORM AN ABLATED TISSUE AREA CONTAINING A PLURALITY OF LESIONS

James B. Hissong and Fred B. Dinger, assignors to Medtronic Xomed, Incorporated
22 July 2003 (Class 601/3); filed 31 July 2000

This device relates to the treatment of anatomical tissue of the head and/or neck with high-intensity focused ultrasound energy, particularly with the aim of skin rejuvenation by thermal stimulation. Skin rejuvenation through the use of ultrasound involves the steps of positioning an ultrasound emitting element adjacent to the external surface of the skin, transmitting the ultrasound energy from the ultrasound element into the skin, focusing the ultrasound energy in the skin, ablating the skin with the focused ultrasound energy to form an ablated tissue area below the external surface of the skin, and removing the ultrasound emitter from the skin. The resultant lesions cause the stimulation of collagen production by the skin. The lesions may begin and end at predetermined depths beneath the external skin surface so that the epidermis and the deep layer of the dermis are not damaged.—DRR



6,599,246

43.80.Sh APPARATUS AND METHOD FOR SUBSTANTIALLY STATIONARY TRANSDUCER THERAPEUTIC ULTRASOUND TREATMENT

Kenneth W. Coffey and Gregory F. Dorholt, both of Tulsa, Oklahoma
29 July 2003 (Class 600/437); filed 25 January 2002

This apparatus provides ultrasound therapeutic treatment during which the transducer is maintained substantially stationary relative to the area being treated. The transducer consists of a plurality of piezoelectric crystals that are spaced apart and sequentially activated for preselected times, thus allowing the transducer to be substantially stationary with respect to the treatment area.—DRR

6,601,581

43.80.Sh METHOD AND DEVICE FOR ULTRASOUND DRUG DELIVERY

Eilaz Babaev, assignor to Advanced Medical Applications, Incorporated
5 August 2003 (Class 128/200.16); filed 1 November 2000

This is a pocket-sized or hand-held ultrasonic oral-pulmonary liquid and powder drug delivery device. In one embodiment, the device is designed

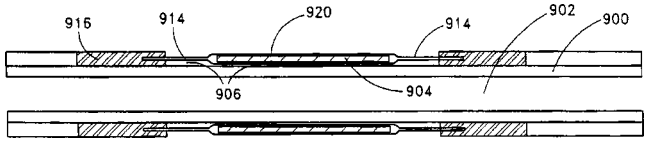
to produce a spray and deliver it to the oral/pulmonary tract without the use of air or other propellants and in such a manner that the spray has about 90% uniform particle size, with a claimed high degree of sharp dosage control. The directed spray is produced by contacting the liquid or powder with a radiation surface of an ultrasonic transducer such as a piezofilm or ultrasound tip. The ultrasonic waves cause the spray to spew outwardly from the radiation surface, directing it into a patient's mouth or to another target.—DRR

6,599,288

43.80.Sh APPARATUS AND METHOD INCORPORATING AN ULTRASOUND TRANSDUCER ONTO A DELIVERY MEMBER

Mark A. Maguire *et al.*, assignors to Atrionix, Incorporated
29 July 2003 (Class 606/27); filed 16 May 2001

This apparatus is a surgical device that provides an ultrasound transducer assembly mounted on a catheter shaft for ultrasonic coupling to a tissue region in a patient and, more specifically, for ultrasonic coupling to a circumferential tissue region in a location where a pulmonary vein extends



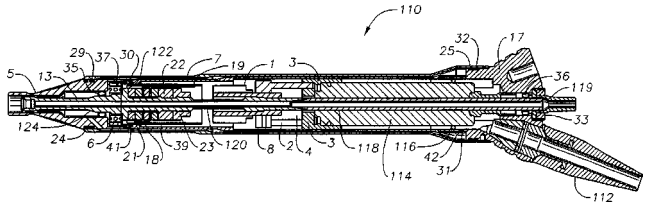
from an atrium. The ultrasound transducer is mounted on support structures that do not bridge the gap between the transducer and the delivery member. The location of the support structures provides for an ''air-backed'' transducer that is claimed to be very efficient and to prevent heat build-up in the materials in contact.—DRR

6,602,193

43.80.Sh ULTRASOUND HANDPIECE

James Y. Chon, assignor to Alcon, Incorporated
5 August 2003 (Class 600/439); filed 22 May 2002

This handpiece, an ultrasonic surgical device intended principally for ophthalmic phacoemulsification, consists of a set of longitudinally vibrating piezoelectric elements and may also include an electric motor to provide rotational or oscillatory movement to the ultrasonic horn. A high-



temperature plastic sleeve, which is filled with high-temperature potting material, surrounds the piezoelectric elements. The patent asserts that the potting material seals the piezoelectric elements against moisture without affecting their performance.—DRR

6,592,520

43.80.Vj INTRAVASCULAR ULTRASOUND IMAGING APPARATUS AND METHOD

Michael Peszynski *et al.*, assignors to Koninklijke Philips Electronics N.V.
15 July 2003 (Class 600/437); filed 31 July 2001

This aparatus includes a catheter with an ultrasound tranducer array. A part of the catheter that can be articulated is connected to a positioning device that allows the array to be placed in a selected orientation relative to the region being examined. The core of the catheter may be rotated or oscillated over an angular range so that ultrasonic image data may be acquired in multiple image planes.—RCW

6,592,523

43.80.Vj COMPUTATIONALLY EFFICIENT NOISE REDUCTION FILTER FOR ENHANCEMENT OF ULTRASOUND IMAGES

Gopal B. Avinash and Pinaki Ghosh, assignors to GE Medical Systems Global Technology Company, LLC
15 July 2003 (Class 600/443); filed 17 July 2002

Noise in an ultrasonic image is reduced by shrinking an initial image and processing the shrunken image with a segmentation-based filtering technique that identifies and differentially processes structures within the image. The segmentation uses a threshold and the distance from the near field of the image. After processing, the shrunken image is enlarged to the dimensions of the initial image and blended with the initial image to produce a final image. During blending, intensity-dependent, random noise is added in nonstructural regions to reduce speckle where the intensity is above a threshold.—RCW

6,592,526

43.80.Vj RESOLUTION ULTRASOUND DEVICES FOR IMAGING AND TREATMENT OF BODY LUMENS

Jay Alan Lenker, Laguna Beach, California
15 July 2003 (Class 600/463); filed 17 October 2000

These devices include a catheter that contains an array of ultrasound transducers and an actuator. The array is oscillated by the actuator to produce more image points than a stationary array. The actuator is solid state and made from nitinol that changes dimension when heated resistively by an electric current. The array may also be moved along the axis of the catheter to image a pyramidal-shaped volume.—RCW

6,595,921

43.80.Vj MEDICAL DIAGNOSTIC ULTRASOUND IMAGING SYSTEM AND METHOD FOR CONSTRUCTING A COMPOSITE ULTRASOUND IMAGE

Joseph A. Urbano and Anthony P. Lannutti, assignors to Acuson Corporation
22 July 2003 (Class 600/437); filed 24 September 2001

A composite ultrasonic image is constructed from visual and audio components that are stored separately. The audio component may be either Doppler sounds or voice. The reconstructed composite image may also be processed in various ways.—RCW

6,595,925

43.80.Vj IN OR RELATING TO CONTRAST AGENTS

Jonny Østensen *et al.*, assignors to Amersham Health AS
22 July 2003 (Class 600/458); filed in the United Kingdom 19 August 1997

The useful time over which gas microbubble contrast agents may be imaged is increased into their recirculating phase in the body by increasing the dosage, imaging at ultrasound frequencies of 2 MHz or less, and by harmonic imaging that uses transmit frequencies less than the resonant frequencies of the microbubbles.—RCW

6,599,248

43.80.Vj METHOD AND APPARATUS FOR ULTRASOUND DIAGNOSTIC IMAGING

Tadashi Tamura, assignor to Aloka
29 July 2003 (Class 600/454); filed 20 March 2001

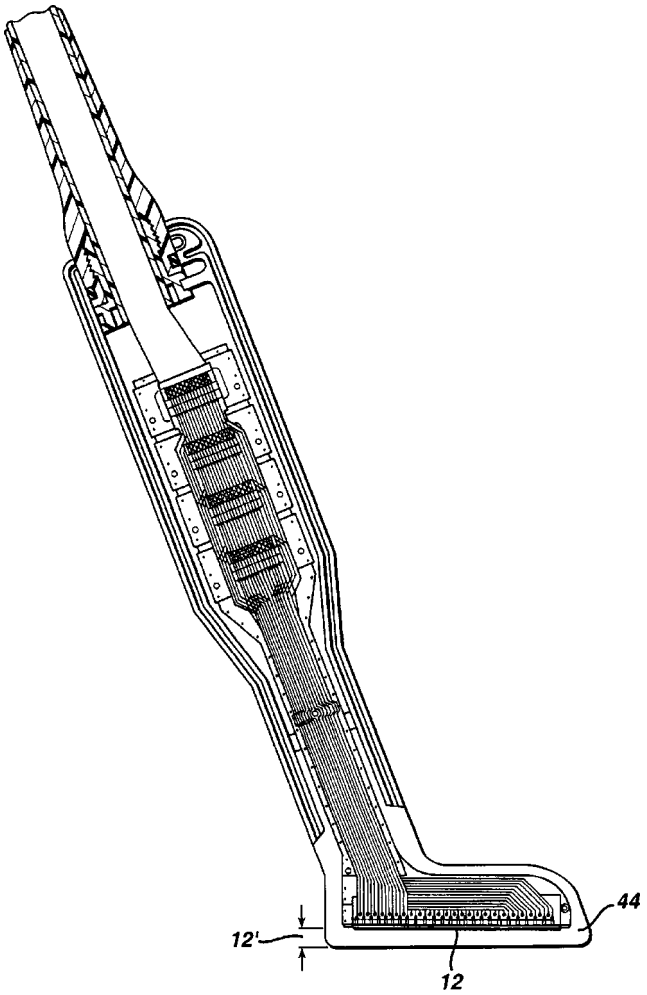
A composite image of tissue and blood flow is produced by combining a beamformed tissue signal and a beamformed blood flow signal. The blood flow signal is obtained by analysis of the difference between a beamformed signal and a delayed or displaced version of the beamformed signal. The beamformed signals may be resampled as well as temporally filtered.—RCW

6,599,249

43.80.Vj INTRAOPERATIVE ULTRASOUND PROBE WITH AN INTEGRATED ACOUSTIC STANDOFF

Timothy Nordgren *et al.*, assignors to Koninklijke Philips Electronics N.V.
29 July 2003 (Class 600/459); filed 14 February 2002

The standoff in this probe is joined to a lens that is in turn joined to the transducer. The focal zone is in the area immediately below the patient contact surface of the probe. The standoff is separated from the surface



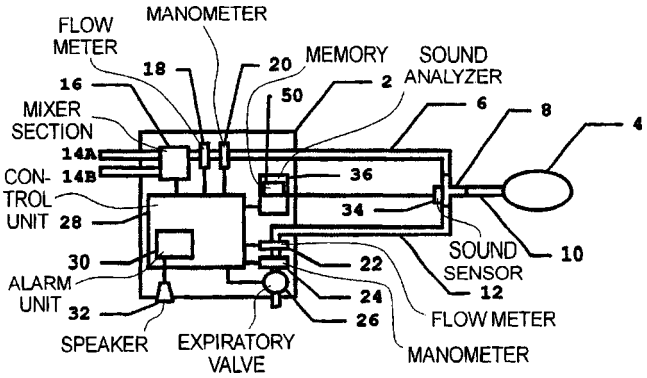
being imaged by a cap comprised of a biocompatible elastomer that enables the probe to be easily sterilized and disinfected as well as providing protection against electrical shock.—RCW

6,601,583

43.80.Vj VENTILATOR WHEREIN OPERATION IS MODIFIABLE DEPENDENT ON PATIENT SOUNDS

Tom Pessala and Mats Cardell, assignors to Siemens Elema AB
5 August 2003 (Class 128/204.23); filed in Sweden 25 January 2000

This is a ventilator for respiratory care. When a patient attempts to speak, a false alarm may be triggered or a patient may be unconscious and unable to trigger an alarm. The sound detector in the ventilator is arranged to



detect and identify sounds that are intentionally made by a patient and sounds made by a patient too weak to press the alarm button in case of pain or discomfort.—DRR

6,602,194

43.80.Vj DUAL BEAMFORMER ULTRASOUND SYSTEM FOR 2D AND 3D IMAGING

David N. Roundhill and Roy B. Peterson, assignors to Koninklijke Philips Electronics N.V.
5 August 2003 (Class 600/443); filed 9 August 2001

This system contains space for the installation of a three-dimensional beamformer that provides the capability to image in three dimensions by use of a scanhead select module.—RCW

6,602,195

43.80.Vj MEDICAL ULTRASONIC IMAGING PULSE TRANSMISSION METHOD

Sriram Krishnan *et al.*, assignors to Acuson Corporation
5 August 2003 (Class 600/447); filed 22 January 2001

This method uses a minimum of three pulses including at least two pulses of different amplitude and at least two pulses of different phase. The larger amplitude pulse is transmitted with a larger aperture and the smaller amplitude pulses are transmitted with respective smaller subapertures. The sum of the subapertures used for the smaller amplitude pulses is equal to the aperture used for the larger amplitude pulse. This allows pulses of differing amplitudes to be obtained without varying the power level of individual transducer element excitations and provides precise control over pulse amplitude.—RCW

6,605,042

43.80.Vj METHOD AND APPARATUS FOR ROTATION REGISTRATION OF EXTENDED FIELD OF VIEW ULTRASOUND IMAGES

Fang F. Dong and Brian Peter Geiser, assignors to GE Medical Systems Global Technology Company, LLC

12 August 2003 (Class 600/447); filed 10 August 2001

An extended field of view image is produced by aligning pixels that correspond to the same spatial points in two images. The alignment is accomplished by rotating the two images. The angle of rotation is calculated using a least-mean-square-error distance between the common pixel points.—RCW

Reply to the Comment on "Multiple scattering in a reflecting cavity: Application to fish scattering" [J. Acoust. Soc. Am. 113, 2978–2979 (2003)] (L)

Julien De Rosny
Laboratoire Ondes et Acoustique, ESPCI, Université Paris VII, CNRS UMR 7587, Paris, France

Philippe Roux
Marine Physical Laboratory, Scripps Institution of Oceanography, UCSD, San Diego, USA

(Received 11 July 2003; revised 11 October 2003; accepted 21 October 2003)

In a recent letter, Ye and Chu comment on a previously published paper by the present authors ["Multiple scattering in a reflecting cavity: Application to fish counting in a tank," J. Acoust. Soc. Am. **109**, 2587–2597 (2001)]. This reply answers the questions asked in that letter and seeks to clarify some of the details in the original paper that may have led to a misunderstanding. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1632481]

PACS numbers: 43.20.El, 43.30.Gv, 43.80.Ev [ADP] Pages: 31–34

The original paper[1] was born from the following problem: how to extract information about a school of fish swimming inside a water tank by means of an acoustic device? To this end, the physics of wave propagation in a system made of a set of moving scatterers (fish, for example) inside a reverberant cavity (such as a tank or an aquarium) had to be understood. The difficulty of the analysis comes from the fact that the wave is multiply scattered during propagation by the moving scatterers as well as multiply reflected by the cavity boundaries. In the original paper, only highly reverberating cavities are considered. When an acoustic pulse travels inside the cavity in the presence of fish, the resulting signal received on a hydrophone is composed of numerous echoes that die out after several hundreds cavity crossings. Among those echoes, some have been reverberated by the cavity interfaces only while others have also been scattered by the fish. The scattered echoes vary of course in amplitude and time with respect to the fish position inside the cavity. As a consequence, if several acoustic pulses are emitted sequentially while the fish are moving inside the cavity, the respective recordings on the hydrophone are composed of both a constant background signal corresponding solely to the reflections on the cavity boundaries and a shot-dependent signal that is related to the fish scattering properties. By averaging signals recorded for different fish positions, the scattered part cancels out and only the part of the field that has never been scattered remains. This acoustic field known as the coherent field is well understood in the frame of multiple scattering theory both in optics[2,3] and acoustic.[4] Indeed, the density and size of the scatterers can be extracted from the measurement of the coherent field. The goal of our work was to extend this theory to the case of moving scatterers inside a reverberating cavity in which multiple scattering is combined with reverberation.

The Comment is composed of four main paragraphs concerning (1) the method of images described in the introduction of the original paper, (2) the mathematical derivation of the theory, (3) the measurement of the total cross section of fish using this method and (4) discussion about the fish counting technique. In the following, we aim to give a convincing explanation to each criticism developed in the Comment. To insure clarity, this Reply will follow as much as possible the order of the main objections raised in the Comment.

(1) The first Comment criticism is based on Fig. 1 of the original paper (also Fig. 1 here), in which "the method of images" is applied to "understand acoustic propagation inside the cavity+scatterers system." The goal of the image method is to transform a complicated problem—acoustic propagation inside a cavity with moving scatterers—into a known problem. To this end, we have applied the well-known image theorem in order to remove the cavity boundaries and replace them by the multiple images of the scatterers and the source. The ray paths in Fig. 1 show that each ray emitted in the cavity from the source can be interpreted as virtual rays emitted in "free space" (i.e., without the cavity) by the images of the source. A scattering event on a scatterer inside the cavity is then replaced by a scattering event on the scatterer image in "free space."

As stated in the original paper, "The cavity+scatterers system is thus equivalent to a sources+scatterers system with no interface." When the image theorem is applied, there is no single source as in the cavity problem, but an infinite number of source images that emit the same pulse simultaneously. The propagation medium is then composed of the scatterers inside the cavity and their images through the cavity interfaces. The signal recorded on the receiver is the sum of the contributions of *all* the image sources propagating through the infinite multi-scattering "image" medium. In other words, the acoustic field received inside the cavity +scatterers system from a unique source is formally equivalent to the signal received at the same point from the images of the source in the infinite multi-scattering "image" medium.

Another criticism of the Comment concerns "the reflection from the water surface" that is not taken into account in the empirical approach described in Fig. 1. Indeed, the authors clearly specified in the original paper that "For sim-

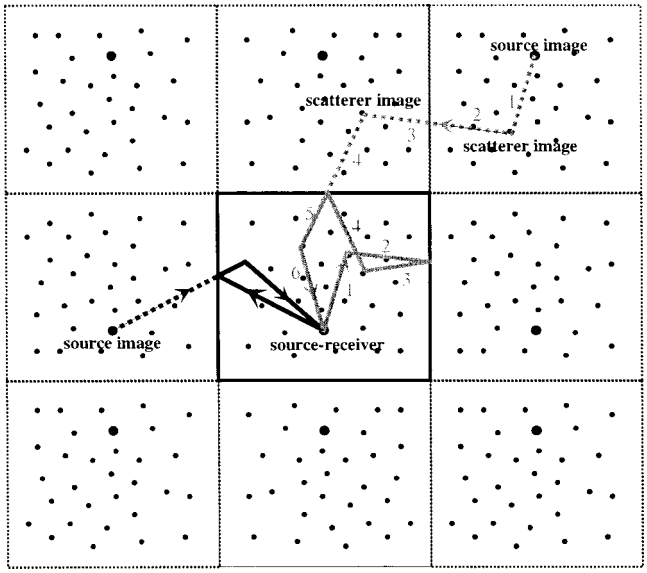


FIG. 1. Representation of two ray paths in a square cavity (black and gray full lines) and their analog in a medium without interface (black and gray dashed lines) using the method of images. When a full-line ray is reflected on the cavity boundary, its ''free space'' analog continues to propagate straight (see segments 2 and 3 inside the cavity and their analog in ''free space''). Each scattering inside the cavity corresponds to a similar scattering on a scatterer image (see segments 1 and 2 in the cavity and their analog in ''free space''). Note that segments 5 and 6 are identical in the two representations. The source-receiver full-line path inside the cavity is then formally equivalent to the dashed path between the source image in ''free space'' and the receiver in the cavity.

plicity, we suppose that the cavity is a 2D square in which scatterers are randomly distributed.'' However, the extension to a three-dimensional or irregularly-shaped cavity is relatively straightforward using the mathematical apparatus developed in Sec. III of the original paper. As a matter of fact, all the experiments reported in the original paper have been performed in 3-D cavities, thus confirming the solid foundation of our theoretical approach.

Off course, the ray picture as depicted in Fig. 1 does not include boundary conditions at the cavity interfaces. To do so, we have to remember that a pressure release interface will transform the original source into a negative source image (with a phase shift of π) whereas a rigid-type boundary generates a positive source image (with a null phase shift). Thus, a positive or negative source image emits a positive or negative pulse, respectively. Attenuation at the interfaces must also be accounted for by weighting the source images according to the reflection coefficients at the cavity interfaces. In Fig. 1, the source image on the left corresponds to a path reflected once inside the cavity and should be weighted by the reflection coefficient at this interface. Similarly, the source image on the upper right image of the cavity corresponds to a path reflected twice on the cavity interfaces. Thus it should be weighted by the product of the reflection coefficients at these interfaces. As a consequence, the number of images of the source in the ''image'' medium that effectively contribute to the signal at the receiver remains finite. The further the source image from the receiver, the smaller (in amplitude) and the later (in time) its contribution to the acoustic field.

To summarize, what is the interest of this ''free space'' representation of the cavity problem? When we wrote the original paper, multiple scattering theory only existed for an opened medium, in the absence of cavity. The concept of coherent and incoherent fields were not properly defined in a system combining multiple scattering and strong reverberation. For example, the existence of a coherent field is not intuitive when a source and a receiver are collocated in a cavity. Therefore, the ''free space'' representation of the cavity problem appeared as an elegant application of the image theorem and the best way to exploit the multi-scattering existing theory that links the coherent field to the mean free path.

In the same paragraph of the Comment, another criticism relies on the fact that ''in a reflecting cavity, the scatterers and their images are *not* independent scatterers.'' The authors of the Comment argue that the field scattered by a scatterer and its image are correlated. They base this assertion on several articles that describe the interaction between a scatterer and a flat interface. The Comment especially refers to Ref. 5. However, Ref. 5 shows that a scatterer and its images are coupled if and only if the scatterer is close to a boundary within less than a few wavelengths. We would like to add that the same effect occurs when two scatterers are close: the resulting field also differs from the independent sum of the waves scattered by the two scatterers. In the original paper, our whole approach is based on a homogeneous and diluted distribution of scatterers in the cavity. These conditions are required in order to relate the mean-free path l_s to the density of scatterers n and their scattering cross section σ_s. It is clearly mentioned in the original paper ''As long as the density of scatterers is not too high, we have the classical equation $l_s = 1/n\sigma_s$.'' That implies that the scatterers are far way from each other (more than a few wavelengths) and they are also far from the reflecting walls. In this case, a coupling between a scatterer and its image is not expected. Nevertheless, even if the cited reference seems for us not contradictory with our results, we agree that the repetition of the same random pattern due to the image theorem (see Fig. 1) may generate some ''nonrandom'' correlations on the field. However, this specific problem has been addressed for more than 50 years in the room acoustic community.[6,7] It has been shown that inside strong reverberant rooms the wave interference that induces correlations between multiple arrivals can be neglected. This standard assumption in room acoustics, so called the diffusion regime, has a good mathematical basis. This approximation is experimentally valid when the pulse bandwidth is wide and the number of reflections on the walls is large. These two conditions are fulfilled in our case: our pulse bandwidth is $\Delta f/f \sim 25\%$ and we only consider late arrival echoes that have been reverberated several times.

In conclusion of this part, we have to keep in mind that, though this is a convenient way to understand acoustic propagation in a cavity, we have to be careful when using the method of images. This is only a phenomenological approach to the problem, and, as such, it is only presented in the introduction of the original paper. It should not be further developed because it is based on a ray picture of acoustic

propagation in a cavity which is obviously not valid in the presence of pointlike scatterers. Moreover, the point-to-point transformation through the cavity interfaces is no longer true in the case of curved or odd-shaped boundaries that are found in all practical tanks or aquariums.

Finally, the authors would like to mention that a recent paper[8] has confirmed all the theoretical results deduced from the phenomenological image theorem of the original paper. Reference 8 is devoted to a more general approach of the reverberant cavity filled with moving scatterers in the framework of the diffusing wave spectroscopy. The theory is based only on a diffusive field assumption and does not invoke the image method.

(2) The second main criticism of the Comment deals with the mathematical formulation of the theory developed in the original paper.[1] Criticisms are based on the very general Eqs. (1)–(4) in the introduction of the paper, while no comment is done on the fully-detailed theoretical approach derived in Sec. III [Eqs. (8)–(12)] and in Appendix A [Eqs. (A1)–(A7)] and B [Eqs. (B1)–(B5)].

The Comment suggests that the definitions of the incoherent and the coherent intensities given by Eqs. (1) and (2) (see below) are incorrect. We do not agree. There is only one definition of the coherent and the incoherent intensities that is totally independent of the disordered medium studied. The incoherent intensity I_t is the squared field averaged over realizations $I_t(t)=\langle h^2(t)\rangle$. The coherent intensity I_c is the square of the averaged field (also called coherent field) $I_c(t)=\langle h(t)\rangle^2$. Here $h(t)$ is the reverberated-scattered signal recorded at the receiver in the cavity after a pulse was emitted from the source, and $\langle\cdot\rangle$ represent the average over the disorder realizations. In our case, different realizations are obtained because of the fish motion between two consecutive pulses.

We invoke now multiple scattering theory between the images of the source and the receiver separated by a distance L.[2–4] As stated in the original paper, ''The coherent and incoherent intensity follow the equations:

$$I_t(L)=I_0\exp(-L/l_a), \tag{1}$$

$$I_c(L)=I_0\exp(-L/l_a)\exp(-L/l_s), \tag{2}$$

where l_s is the scattering mean free path, l_a is the absorption mean free path, and I_0 is the field intensity.'' The scattering mean free path l_s corresponds to the part of the field that has been elastically scattered by the fish and then lost in the average of the coherent field. The attenuation l_a englobes attenuation in the medium, loss at the interfaces, and inelastic scattering on the scatterers. We observe then that the ratio

$$\frac{I_c(L)}{I_t(L)}=\exp\left(-\frac{L}{l_s}\right) \tag{3}$$

does not depend on the source excitation. As written in the original paper ''This ratio only depends on the elastic scattering of the scatterers whatever the attenuation in the medium.'' These statements show that the attenuation of the source images due to the reflection coefficients at the cavity interfaces [see paragraph (1) above] does not appear in Eq. (3). Only the distance L between the source images and the receiver matters.

In the same paragraph of the Comment, the transition from the space variable in Eq. (3) to the time variable in Eq. (4) is not understood. The Comment stated that Eq. (3) is only valid in ''a *forward* propagating process'' while the cavity problem is ''a combination of the forward process, via reflections, and the *backscattering* process.'' This reveals a misconception of the multi-scattering problem in a cavity and brings us back to the interpretation of Fig. 1 and the theorem of images. We first recall Eq. (4):

$$\frac{I_c(t)}{I_t(t)}=\exp\left(-\frac{t}{\tau_s}\right). \tag{4}$$

We state in the original paper that ''$\tau_s=l_s/c$, c is the sound speed in the medium, and time t is the time arrival of the coherent echo between each source image and the actual receiver in the cavity.'' As stated in paragraph (1), the distance L between a source image and the receiver is equivalent to the acoustic path of the corresponding reverberated echo inside the cavity. Thus, an echo of the coherent field traveling a net distance L will arrive after a time interval $t=L/c$. The keypoint here is that coherent signals travel along piecewise straight paths between each source image and the receiver at the speed of sound.

However, one may wonder in which case the presence of scatterers in the medium will affect the relation $L=ct$. In our problem, there is no ambiguity because of the diluted-medium approximation [see paragraph (3) below for further details]. In the case of a dense cloud of small scatterers, like plankton, for example, the density of scatterers may affect the sound speed in the medium, which implies a rescaling of the sound speed c as described in Ref. 9.

(3) The last criticism deals with the absolute measurement of the total scattering cross section of the scatterers. The Comment correctly pointed out that the value of the sphere radius $R\sim 3.8$ cm whose scattering cross section is equivalent to the cross section of a 35-cm-long striped bass is not realistic. In the original paper, the measurement of the target strength TS$=-35$ dB is correct. It is simply deduced from the slope of the experimental curve displayed in Fig. 16. From the definition of TS, TS$=10\log_{10}(\sigma_s/4\pi)$, σ_s can be easily obtained and is roughly equal to 40 cm^2. The numerical error only appears when the equivalent-sphere radius R is computed. Indeed we have $\sigma_s=2\pi R^2$ (Ref. 10) (and not $\sigma_s=4\pi R^2$ as in the original paper) from which the correct value $R\sim 2.5$ cm is deduced. We agree that the value of R of the original paper was confusing. Nevertheless, the value of TS (and then of σ_s) given in the original paper is in good agreement with the values found in the literature cited in the Comment.[11] In the same part, the Comment gives a detailed description of the total cross section as the sum of the total scattering and absorption cross sections. This is completely irrelevant to the original paper in which only the total scattering cross section is evaluated. To avoid further criticism, we would like also to mention that though the formula of the target strength (TS) given above is only valid for the case of

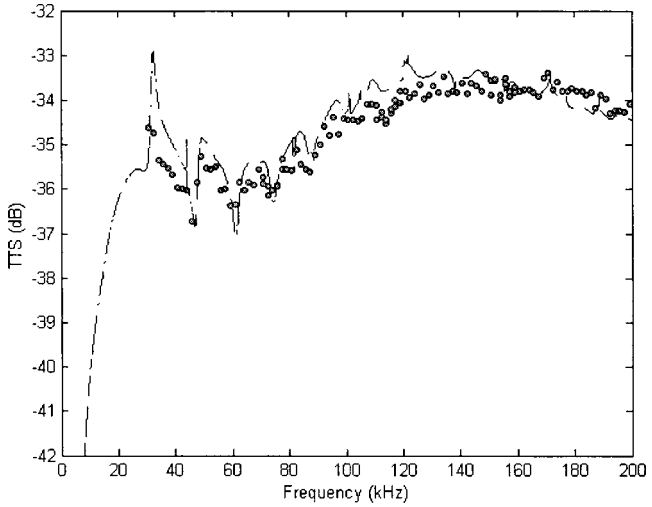


FIG. 2. Comparison between the frequency-dependent total target strength (TTS) obtained theoretically (dashed line) and experimentally using our acoustic reverberation technique (filled circles) for a 60-mm-diam copper sphere. More precision can be found in Ref. 12 about the experimental setup.

an omni-directional scatterer, it is widely used in fisheries acoustics to give an acoustic index of reflection for fish.

Two other results in the original paper confirm that the measurement of the mean-free path l_s provides an accurate estimate of the total scattering cross section σ_s. In Sec. II, numerical results obtained in a 2-D cavity with scatterers using a finite difference time domain simulation (Fig. 4) show that σ_s is retrieved within 2% accuracy. In Sec. IV, experiments performed at 400 kHz in a beaker with 1-cm-long zebra fish show a perfect linearity between the measurement of $1/l_s$ and the number of fish, in complete agreement with Eq. (5) ($l_s = 1/n\sigma_s$). This shows that σ_s does not depend on the number of fish in the beaker. Similarly, experimental results in Fig. 15 prove that σ_s is independent on the volume of the beaker. In conclusion of this set of ultrasonic experiments, the scattering cross section of 1-cm-long zebra fish is found within 3% accuracy. This study confirms that the above technique can be efficiently used as a fish counting technique or as a fish sizing technique.

If any other proof was to be mentioned, an interesting demonstration of the measurement accuracy of the total scattering cross section using our reverberation technique is published in Ref. 12. In this paper a simple practical application of Eqs. (1)–(4) is performed using a calibrated sphere moved manually inside a reverberating cavity. The results show that the theoretical and experimental frequency-dependent total scattering cross sections of the sphere are in perfect agreement within 0.5 dB (Fig. 2).

Finally, we have recently extended the application of this reverberation technique to the measurement of the absorption cross section of scatterers.[13] This may be of interest to quantitatively study the relative proportion of acoustic scattering and absorption for complex scatterers like fish.

(4) We would like to thank the authors of the Comment for recognizing that "the approach in the paper tends to get good fish number counts, a promising aspect of the method." Actually, as seen above, this technique cannot provide a good fish-count without a good fish-size estimate, the two quantities being related via the scattering mean free path l_s.

To summarize, this reverberation-based method enables the measurement of the mean free path l_s of the scatterers in the cavity—leading to the measurement of the number of scatterers and/or to the total scattering cross-section σ_s—even in the case where l_s is much larger than the dimensions of the cavity. In other words, the use of a reverberant cavity makes it possible to measure σ_s for a weak scatterer because the scatterer is encountered not only once by the acoustic wave, but multiple times and from multiple directions thanks to the reverberation in the cavity.

In conclusion, the authors replied to all the criticisms raised in the Comment published on their work. We hope to have clarified some details in the original paper. According to us, both the earlier experimental results (beside a numerical error) and our theoretical approach remain correct.

ACKNOWLEDGMENT

The authors would like to thank Stephane Conti and Dave A. Demer for their contribution to this work.

[1] J. De Rosny and P. Roux, "Multiple scattering in a reflecting cavity: Application to fish counting in a tank," J. Acoust. Soc. Am. **109**, 2587–2597 (2001).
[2] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Oxford U. P., Oxford, 1997), p. 253.
[3] P. Sheng, *Introduction to Wave Scattering, Localization, and Mesoscopic Phenomena* (Academic, San Diego, 1995), Chaps. 3 and 4.
[4] A. Derode, A. Tourin, and M. Fink, "Random multiple scattering of ultrasound I. Coherent and ballistic wave," Phys. Rev. E **64**, 036606 (2001).
[5] I. Tolstoy, "Super-resonant systems of scatterers. I," J. Acoust. Soc. Am. **80**, 282–294 (1986).
[6] E. Meyer, "Definition and diffusion in rooms," J. Acoust. Soc. Am. **26**, 630–636 (1954).
[7] M. R. Schroeder, "Measurement of sound diffusion in reverberation chambers," J. Acoust. Soc. Am. **31**, 1407–1414 (1959).
[8] J. De Rosny, P. Roux, M. Fink, and J. H. Page, "Field fluctuation spectroscopy in a reverberant cavity with moving scatterers," Phys. Rev. Lett. **90**(9), 094302 (2003).
[9] D. Chu, P. Wiebe, and N. Copley, "Inference of material properties of zooplankton from acoustic and resistivity measurements," ICES J. Mar. Sci. **57**, 1128–1142 (2000).
[10] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton U.P., Princeton, NJ, 1968), pp. 419–420.
[11] K. G. Foote, "Analysis of empirical observations on the scattering of sound by encaged aggregations of fish," Fiskeridir. Skr., Ser. Havunders. **16**, 422–455 (1978).
[12] D. A. Demer, S. Conti, J. De Rosny, and P. Roux, "Absolute measurement of total target strength from reverberation in a cavity," J. Acoust. Soc. Am. **113**, 1387–1394 (2003).
[13] S. Conti, P. Roux, D. A. Demer, and J. De Rosny, "Measurement of the scattering and absorption cross-sections of the human body," Appl. Phys. Lett. (in press, 2004).

A simplified algorithm for second-order sound beams with arbitrary source distribution and geometry (L)

Desheng Ding
Department of Electronic Engineering, Southeast University, Nanjing 210096, People's Republic of China

(Received 11 August 2003; revised 29 October 2003; accepted 31 October 2003)

An earlier investigation [J. Acoust. Soc. Am. **108**, 2759–2764 (2000)] of the simplified algorithm for the second-order sound fields with circular axial symmetry is extended to a more general case of the sound sources with an arbitrary distribution. This extension is based on the fact that the source function is expanded into the sum of a series of two-dimensional Gaussian ones. The two- and five-dimensional integral expressions for the primary and second-order fields are, respectively, reduced in terms of Gaussian functions and simple one-dimensional integrals. Numerical calculations are then greatly simplified. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1635409]

PACS numbers: 43.25.Lj, 43.25.Cb [MFH] Pages: 35–37

There has been steady growth of interest in the nonlinear propagation of bounded sound beams in many fields such as acoustic microscopy, medical ultrasound, parametric acoustic arrays, and nondestructive testing.[1–3] When the sound-pressure level is moderate and all the harmonics that are higher than the second order are assumed to be negligible; that is, under the quasilinear approximation, the primary (fundamental) and second-order sound beams (second-harmonic, sum-frequency, and difference-frequency generation) are derived in integral expressions by using the method of successive approximations.[4,5] These integrals are strongly oscillatory and multiple dimensional. In particular, the dimension number is *high up to 5* for the second-order sound field. Fortunately, a base function expansion technique provides an alternate approach to greatly simplify the field integrals.[6–13]

In this Letter, we present a more general approach to analysis of the primary and second-order beams. Any source function is expanded into the superposition of a set of two-dimensional Gaussian functions.[12] Accordingly, the primary radiation field is expressed as a linear superposition of this set of Gaussian beams and the second-order generation is then considered as a sum of the interaction terms by these Gaussian ones. Eventually, the two- and five-dimensional integral representations for the primary and second-order field components simplify in terms of Gaussian functions and simple one-dimensional integrals, respectively. The present extension does not require the condition that the source distribution is of the circular axial symmetry and it included the previous results as special cases.[6–10,12]

Parallel to the previous paper,[10] we express the field integrals in terms of dimensionless variables. Assume that two sound sources (or one source excited by different frequency components) with the arbitrary distribution oscillate harmonically at two different angular frequencies ω_j ($j=1,2$ and $\omega_1>\omega_2$), and are located in the (x,y) plane normal to the z axis, the direction of the beam propagation. We formally have the primary sound fields[1,4,5,12,13]

$$p_1^{(j)}(\xi_j,\zeta_j,\eta_j;\tau_j)=\mathrm{Re}[p_{0j}e^{-i\tau_j}\overline{q}_1^{(j)}(\xi_j,\zeta_j,\eta_j)], \tag{1a}$$

where

$$\overline{q}_1^{(j)}(\xi_j,\zeta_j,\eta_j)=\frac{1}{i\pi\eta_j}e^{-\overline{\alpha}_j\eta}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left(i\frac{(\xi_j-\xi_j')^2+(\zeta-\zeta_j')^2}{\eta_j}\right)\times\overline{q}_1^{(j)}(\xi_j',\zeta_j')d\xi_j'd\zeta_j', \tag{1b}$$

and the second-order sounds fields

$$p_{lm}(\xi,\zeta,\eta;\tau)=\mathrm{Re}\left\{-(l+m)^2p_{01}p_{02}\frac{\beta k^2S}{\rho c^2}e^{-i(l+m)\tau}\overline{q}_{lm}(\xi,\zeta,\eta)\right\}, \tag{2a}$$

where

$$\overline{q}_{lm}(\xi,\zeta,\eta)=\frac{1}{4\pi}e^{-\overline{\alpha}_\pm\eta}\int_{\eta'=0}^{\eta}\int_{\xi'=-\infty}^{\infty}\int_{\xi'=-\infty}^{\infty}e^{\overline{\alpha}_\pm\eta'}\times\frac{1}{\eta-\eta'}\exp\left(i(l+m)\frac{(\xi-\xi')^2+(\zeta-\zeta')^2}{\eta-\eta'}\right)\times\overline{q}_1^{(1)}(\xi_1',\zeta_1',\eta_1')\overline{q}_1^{(2)}(\xi_2',\zeta_2',\eta_2')d\xi'\,d\zeta'\,d\eta'. \tag{2b}$$

In these equations, $\tau_j=\omega_jt-k_jz$, $p_{0j}=\rho u_{0j}c$ is the characteristic pressure amplitude of primary waves, and u_{0j} is the amplitude of vibration velocity on the source. $k_j=\omega_j/c$ ($k_1>k_2$) denotes the wave number of primaries. ρ, c, and β are the density, sound speed, and acoustic nonlinear parameter of the medium, respectively. Here, $\xi_j=x/\sqrt{S}$, $\zeta_j=y/\sqrt{S}$, and $\eta_j=2z/k_jS$ are transversely and axially dimensionless coordinates, or more properly as auxiliary variables. S is a quantity simply related to the source area. For example, it may be defined as $S=ab$ for an elliptical piston with the semimajor and semiminor axes a and b. The source functions $\overline{q}_1^{(1)}(\xi_1',\zeta_1')$ and $\overline{q}_1^{(2)}(\xi_2',\zeta_2')$ may have the arbitrary distribution. In Eq. (2), the other notations are defined in such a way: $\eta=2z/kS$, $k=(k_1+k_2)/2$, $l=k_1/k$, $m=k_2/k$, and $\tau=\omega t-kz$ with $\omega=kc$. Additionally, α_1, α_2, α_+, and α_- are the

small signal absorption coefficients at the angular frequencies ω_1, ω_2, $\omega_1+\omega_2$, and $\omega_1-\omega_2$, respectively; these notations with the bar represent the attenuation loss at the range $z_0=\frac{1}{2}kS$. Obviously, when $l=m=1$, Eq. (2b) is just the expression for the second-harmonic generation. With the exception for this case, Eq. (2a) should be multiplied by the factor 1/2. If we simply take $m=-k_2/k$, a negative value, Eq. (2) indeed describes the difference-frequency generation of sound beams. Equations (1b) and (2b) are the complex-valued pressure amplitudes in dimensionless form. Without the loss of generality, in what follows we shall neglect the exponential attenuation terms with $\overline{\alpha}$ in Eqs. (1b) and (2b).

A two-dimensional Gaussian source with the center (a,b) has its distribution function G_2

$$G_2(\xi',\zeta';B_x,B_y)=\exp[-[B_x(\xi'-a)^2+B_y(\zeta'-b)^2]], \quad (3)$$

where B_x and B_y are generally complex numbers with the real part greater than zero. This Gaussian beam field is expressed as[12]

$$G_2(\xi-a,\zeta-b,\eta;B_x,B_y) = G_1(\xi-a,\eta;B_x)\cdot G_1(\zeta-b,\eta;B_y), \quad (4)$$

with G_1

$$G_1(\xi,\eta;B_x)=\frac{1}{\sqrt{1+iB_x\eta}}\exp\left(-\frac{B_x\xi^2}{1+iB_x\eta}\right). \quad (5)$$

In the derivation, a well-known formula

$$\int_{-\infty}^{\infty}\exp\left(-\frac{t^2}{4\beta}-\gamma t\right)dt=2\sqrt{\pi\beta}\exp(\beta\gamma^2), \quad (6)$$

is used.

It is assumed that the source function $\overline{q}_1^{(j)}(\xi_j',\zeta_j')$ with an arbitrary geometry and distribution has the Gaussian expansion[12]

$$\overline{q}_1^{(j)}(\xi_j',\zeta_j')=\sum_{k=1}^{N_j}A_k^{(j)}G_2(\xi_j'-a_k^{(j)},\zeta_j'-b_k^{(j)};B_{x,k}^{(j)},B_{y,k}^{(j)}). \quad (7)$$

Substituting Eq. (8) into (1b) and using Eq. (4), one obtains the primary field

$$\overline{q}_1^{(j)}(\xi_j,\zeta_j,\eta_j) = \sum_{k=1}^{N_j}A_k^{(j)}G_2(\xi_j-a_k^{(j)},\zeta_j-b_k^{(j)},\eta_j;B_{x,k}^{(j)},B_{y,k}^{(j)}). \quad (8)$$

The two-dimensional integral solution of Eq. (1b) reduces to a sum of simple Gaussian functions.

We next give the calculation procedure for the second-order sound beams. Substitution of Eq. (8) into Eq. (2b) yields

$$\overline{q}_{lm}(\xi,\zeta,\eta)=\sum_{k=1}^{N_1}\sum_{j=1}^{N_2}A_k^{(1)}A_j^{(2)}G_4(\xi,\zeta,\eta;k,j), \quad (9)$$

where the function G_4 is of the form

$$G_4(\xi,\zeta,\eta;k,j)=\frac{1}{4\pi}\int_{\eta'=0}^{\eta}\int_{\xi'=-\infty}^{\infty}\int_{\zeta'=-\infty}^{\infty}\frac{1}{\eta-\eta'}\exp\left(i(l+m)\frac{(\xi-\xi')^2+(\zeta-\zeta')^2}{\eta-\eta'}\right)\times G_2(\xi_1'-a_k^{(1)},\zeta_1'-b_k^{(1)},\eta_1';B_{x,k}^{(1)},B_{y,k}^{(1)}) \times G_2(\xi_2'-a_j^{(2)},\zeta_2'-b_j^{(2)},\eta_2';B_{x,j}^{(2)},B_{y,j}^{(2)})d\xi'\,d\zeta'\,d\eta'. \quad (10)$$

This is still a very complicated three-dimensional integral. Again using formula (6) and separately integrating over ξ' and ζ', and letting

$$B_{xl,k}=B_{x,k}^{(1)}/l, \quad B_{yl,k}=B_{y,k}^{(1)}/l, \quad B_{xm,j}=B_{x,j}^{(2)}/m, \quad B_{ym,j}=B_{y,j}^{(2)}/m, \quad (11)$$

finally we have

$$G_4(\xi,\zeta,\eta;k,j) = \frac{1}{4}\int_{\eta'=0}^{\eta}\frac{1}{(r_1\eta'+r_2)^{1/2}}\exp\left(-\frac{s_2(\xi-a_{k,j})^2}{r_1(r_1\eta'+r_2)}-\frac{s_1}{r_1} \times(\xi-a_{k,j})^2-r_4'\right)\frac{1}{(q_1\eta'+q_2)^{1/2}} \times\exp\left(-\frac{t_2(\zeta-b_{k,j})^2}{q_1(q_1\eta'+q_2)}-\frac{t_1}{q_1}(\zeta-b_{k,j})^2-q_4'\right)d\eta', \quad (12)$$

where

$$r_1=(lB_{xm,j}+mB_{xl,k})+i(l+m)\eta B_{xl,k}B_{xm,j}, \quad (13a)$$

$$r_2(lB_{xl,k}+mB_{xm,j})\eta-i(l+m), \quad (13b)$$

$$s_1=(l+m)^2B_{xl,k}B_{xm,j}, \quad (13c)$$

$$s_2=-i(l+m)lm(B_{xl,k}-B_{xm,j})^2, \quad (13d)$$

$$r_1'=(lB_{xl,k}+mB_{xm,j})+i(l+m)\eta' B_{xl,k}B_{xm,j}, \quad (13e)$$

$$r_3'=(lB_{xl,k}a_k^{(1)}+mB_{xm,j}a_j^{(2)}) + iB_{xl,k}B_{xm,j}\eta'(la_k^{(1)}+ma_j^{(2)}), \quad (13f)$$

$$a_{k,j}=r_3'/r_1', \quad (13g)$$

$$r_4'=lmB_{xl,k}B_{xm,j}(a_k^{(1)}-a_j^{(2)})^2/r_1', \quad (13h)$$

and the others, q_1, q_2, t_1 and t_2,..., correspond one-to-one to r_1, r_2, s_1, and s_2,... . The difference is that the subscript x of Bs is replaced by y, and a by b. In quasilinear theory, Eq. (12) may be interpreted as the second-order component generated by nonlinear interaction of two Gaussian beams.[3–5,9,10]

It is found that once the set of expansion coefficients from Eq. (7) is known, the result from the present procedure

with use of Eqs. (9), (12), and (13), is considerably simpler to evaluate than the original field integral of Eq. (2b) for the second-order field, requiring at most numerical integration of one dimension. For a given source function, the two-dimensional Gaussian expansion parameters may, in principle, be determined by computer optimization.[7] In some cases, such as for the rectangular and elliptical sources excited uniformly, these parameters are obtained directly from a known result by Wen and Breazeale, by simple combination of their coefficients (Table I of Ref. 7). The details can be found in Ref. 12. For these regular sources, all $a_k=0$ and $b_k=0$ in the expansion parameter, the estimation of the second-order fields on the beam axis (at $\xi=0$ and $\zeta=0$) becomes simpler, involving only in the logarithm functions. In the above treatment, we ignore the attenuation effect of sound. It is easy to recover this effect by adding the exponentials with $\bar{\alpha}$'s into the final formulas (8) and (12). In addition, it is worth pointing out that the present method is easy to apply to the weakly focused case of the second-order sound beams by spherical or cylindrical lenses.[8–10,12]

In conclusion, we have presented a general approach that treats efficiently the second-order sound fields. The complicated field integrals for the primary and second-order components are simply expressed in terms of two-dimensional Gaussian functions and one-dimensional integrals. A reduction of computation amount for the sound beam distributions is expected.

ACKNOWLEDGMENT

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

[1] M. F. Hamilton and D. T. Blackstock, *Nonlinear Acoustics* (Academic, San Diego, 1998).
[2] S. I. Aanonsen, T. Barkve, J. Naze Tøjtta, and S. Tøjtta, "Distortion and harmonic generation in the near field of a finite amplitude sound beam," J. Acoust. Soc. Am. **75**, 749–768 (1984).
[3] M. F. Hamilton, "Effects of noncollinear interaction on parametric arrays in dispersive fluids," J. Acoust. Soc. Am. **76**, 1493–1504 (1984).
[4] B. G. Lucas, J. Naze Tøjtta, and S. Tøjtta, "Field of a parametric focusing source," J. Acoust. Soc. Am. **73**, 1966–1971 (1983).
[5] J. B. Berntsen, J. Naze Tøjtta, and S. Tøjtta, "Near field of a large acoustic transducer. IV. Second harmonic and sum frequency radiation," J. Acoust. Soc. Am. **75**, 1383–1391 (1984).
[6] E. Cavanagh and B. D. Cook, "Gaussian–Laguerre description of ultrasonic fields—Numerical example: Circular piston," J. Acoust. Soc. Am. **67**, 1136–1140 (1980).
[7] J. J. Wen and M. A. Breazeale, "A diffraction beam field expressed as the superposition of Gaussian beams," J. Acoust. Soc. Am. **83**, 1752–1756 (1988).
[8] R. B. Thompson, T. A. Gray, J. H. Rose, V. G. Kogan, and E. F. Lopes, "The radiation of elliptical and bicylindrically focused piston transducers," J. Acoust. Soc. Am. **82**, 1818–1828 (1987).
[9] D. Ding, Y. Shui, J. Lin, and D. Zhang, "A simple calculation approach for the second harmonic sound field generated by an arbitrary axial-symmetric source," J. Acoust. Soc. Am. **100**, 727–733 (1996).
[10] D. Ding, "A simplified algorithm for the second-order sound fields," J. Acoust. Soc. Am. **108**, 2759–2764 (2000).
[11] M. Spies, "Transducer field modeling in anisotropic media by superposition of Gaussian base function," J. Acoust. Soc. Am. **105**, 633–638 (1999).
[12] Y. Zhang, J. Liu, and D. Ding, "Sound field calculations of elliptical pistons by the superposition of two-dimensional Gaussian beams," Chin. Phys. Lett. **19**, 1825–1827 (2002); D. Ding, Y. Zhang, and J. Liu, "Some extensions of the Gaussian beam expansion: Radiation fields of the rectangular and the elliptical transducer," J. Acoust. Soc. Am. **113**, 3043–3048 (2003).
[13] F. H. Fenlon and F. S. McKendree, "Axisymmetric parametric radiation—A weak interaction model," J. Acoust. Soc. Am. **66**, 534–547 (1979).

Effect of talker and speaking style on the Speech Transmission Index (L)

Sander J. van Wijngaarden[a] and Tammo Houtgast
TNO Human Factors, PO Box 23, 3769 ZG Soesterberg, The Netherlands

(Received 5 March 2003; revised 17 October 2003; accepted 30 October 2003)

The Speech Transmission Index (STI) is routinely applied for predicting the intelligibility of messages (sentences) in noise and reverberation. Despite clear evidence that the STI is capable of doing so accurately, recent results indicate that the STI sometimes underestimates the effect of reverberation on sentence intelligibility. To investigate the influence of talker and speaking style, the Speech Reception Threshold in noise and reverberation was measured for three talkers, differing in clarity of articulation and speaking style. For very clear speech, the standard STI yields accurate results. For more conversational speech by an untrained talker, the effect of reverberation is underestimated. Measurements of the envelope spectrum reveal that conversational speech has relatively stronger contributions by higher ($>$ 12.5 Hz) modulation frequencies. By modifying the STI calculation procedure to include modulations in the range 12.5–31.5 Hz, better results are obtained for conversational speech. A speaking-style-dependent choice for the STI modulation frequency range is proposed. [DOI: 10.1121/1.1635411]

PACS numbers: 43.71.Gv, 43.71.Hw [KWG] Pages: 38–41

I. INTRODUCTION

The Speech Transmission Index (IEC, 1998; Steeneken and Houtgast, 1980) is a physical measure for objectively predicting the intelligibility of speech. The Speech Transmission Index (STI) model uses modulation transfer functions (MTFs) to predict intelligibility under influence of a wide diversity of speech degradations, including degradations of a temporal nature, such as reverberation and echoes (Houtgast *et al.*, 1980). Through the modulation transfer function, these influences are translated into ''equivalent speech-to-noise ratios,'' and then treated in essentially the same way as additive noise.

The STI method was designed and optimized to yield representative and homogeneous intelligibility predictions across all kinds of speech degradation, including noise and reverberation. This was validated using consonant–vowel–consonant (CVC) words (Steeneken and Houtgast, 1980), and also found to be true for short, redundant sentences (Duquesnoy and Plomp, 1980).

However, recent experiences with SRT sentences based on more conversational speech (van Wijngaarden *et al.*, 2001) indicate a tendency for the STI to underestimate the effect of reverberation on sentence intelligibility. Indications for a mismatch between subjective intelligibility and the STI in combined ''noise plus reverberation'' conditions are also found in other studies (Payton *et al.*, 1994; Fig. 10, triangular data points on the left). The mismatch reported by Payton *et al.* (1994) seems to depend on speaking style, and is larger for a conversational than for a clear speaking style. This could suggest that the difference may be due to differences in speaking style.

In the next section, experiments along the lines of Duquesnoy and Plomp (1980) are described, in which the effect of noise and (simulated) reverberation on the STI is studied for talkers differing in speaking style.

The version of the STI method used throughout this letter is the revised STI (STI_r), based on the most recent version of the standard available at the time this study was carried out (IEC, 1998).

II. SENTENCE INTELLIGIBILITY IN NOISE AND REVERBERATION

A. Method

The speech reception threshold (SRT; Plomp and Mimpen, 1979) is the speech-to-noise ratio at which 50% intelligibility of short, redundant sentences is realized. The original corpus of speech recordings made by Plomp and Mimpen has seen extensive application, and was also included in the present study.

A new, much larger, corpus of SRT test sentences is the ''VU'' corpus (Versfeld *et al.*, 2000). The sentences by the male talker of the VU corpus were used in this experiment. Versfeld *et al.* present the VU sentences as roughly equivalent to the Plomp and Mimpen sentences. However, the authors of this article perceive the adopted speaking style to be less clear.

A third corpus of SRT sentences is the multilingual SRT (ML-SRT) database (van Wijngaarden *et al.*, 2001; 2002). This corpus consists of material by many nonprofessional talkers in various languages. The single male Dutch talker used in the present study speaks less clearly than the VU talker, and certainly less clearly than the Plomp and Mimpen talker.

The masking noise used in the SRT procedure was noise with the same long-term spectrum as speech by the corresponding talker. Noise was mixed with the target speech samples, after which this signal was convolved with suitable (synthetic) impulse responses to recreate reverberant speech

[a]Electronic mail: vanwijngaarden@tm.tno.nl

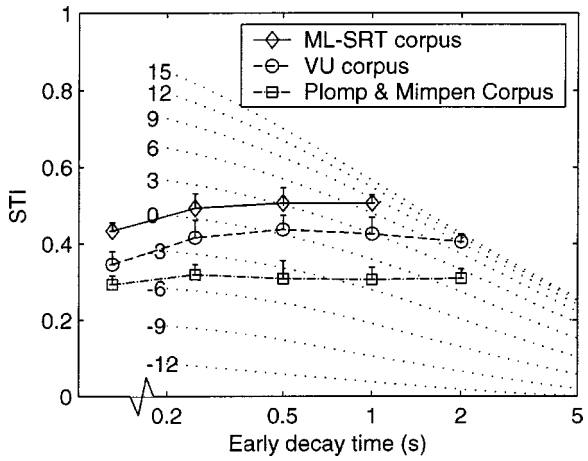


FIG. 1. STI at the SRT (native Dutch listeners), for conditions with and without synthetic reverberation. The dotted lines indicate the maximum STI at each EDT, as a function of the SNR. The error bars indicate the standard deviation (8 listeners, each 2 SRT measurements per condition). The leftmost data point of each line represents a condition without reverberation.

in noise. Use was made of synthetic instead of real impulse response, to exclude effects of nonsystematic differences in timbre for different early decay times (EDTs). In terms of the modulation transfer function, the resulting pseudo-reverberant conditions are identical to purely exponentially decaying naturally reverberant conditions of the same EDT.

B. Results

Duquesnoy and Plomp (1980) measured the SRT as a function of EDT, and then evaluated the STI at this speech-to-noise ratio (the ''STI at the SRT''). They found that 50% sentence intelligibility always corresponds to the same STI, whether noise is the predominant speech degrading factor or reverberation. This experiment was essentially repeated, this time using not only the Plomp and Mimpen (1979) talker used by Duquesnoy and Plomp, but also the more conversational speech taken from the two other corpora described above.

Figure 1 shows STI at the SRT results for the three different talkers. First, the individual SRT was measured in a number of reverberation conditions. From this, the STI at the SRT (the STI corresponding to 50% sentence intelligibility) was calculated. If the STI model predicts effects of reverberation as accurately and unbiased as effects of noise (when related to sentence intelligibility), then the lines in Fig. 1 must be straight and horizontal.

The three talkers represented in Fig. 1 differ in terms of their average intelligibility; the three lines differ significantly. Figure 1 also clearly shows that 50% sentence intelligibility sometimes corresponds to a higher STI *with* than *without* reverberation. For the Plomp and Mimpen talker, the line in Fig. 1 follows the theoretical straight and horizontal line. There is a significant difference only between the STI without reverberation and the STI at EDT$=0.25$ s, but this difference is relatively small. This essentially replicates the results found by Duquesnoy and Plomp (1980). For the VU and ML-SRT talkers, there is a mismatch; the STI without reverberation differs significantly ($p<0.05$) from the STI in any reverberation condition.

III. EXPLANATION FOR THE EFFECT OF SPEAKING STYLE

A. Trends observed in the data

For the upper two lines in Fig. 1, the STI at the SRT is clearly higher at a (relatively small) EDT of 0.2 s than in the absence of reverberation. This implies that even a small amount of reverberation may have an impact on intelligibility, to a degree not predicted by the STI model. This effect only appears for the two talkers adopting a more informal, conversational speaking style.

These observations can be explained by assuming that the relation between the envelope spectrum of speech and intelligibility depends on speaking style. The way that the STI model relates intelligibility to the modulation transfer function is apparently quite suitable for some talkers (and speech styles), but less so for others.

B. Between-corpus differences in the speech envelope spectrum

The STI model uses a fixed (logarithmic) set of 14 modulation frequencies ranging from 0.63 to 12.5 Hz, at 1/3-octave intervals. This represents, more or less, the modulation frequency range observed in natural speech. The envelope spectrum of speech normally shows a maximum around 3 Hz, and contains almost all of its energy in the range from 0–30 Hz.

The modulation frequency range in the STI model, and the choice to give each modulation frequency equal weight, are design choices, optimized to make the STI equally sensitive to all sorts of degradations in time and frequency domain. The chosen range was shown to be appropriate for CVC nonsense words (Steeneken and Houtgast, 1980). As shown above, this validation sometimes holds for short sentences, but apparently only for clear speech by a trained talker. If differences in clarity of articulation and speaking style translate into differences in the envelope spectrum, something may be said for adopting different modulation frequency weighting schemes for different speaking styles.

Envelope spectra were calculated from the recorded SRT sentences. The method for calculating envelope spectra essentially follows the procedure originally proposed in the context of the STI model (Houtgast *et al.*, 1980), but is implemented in digital algorithms rather than analog hardware. The speech (sampled at 44 100 Hz) is band filtered into the seven audio-frequency octave bands used by the STI model. Next, the modulation spectrum is derived from the squared signal by means of a discrete Fourier transform. The obtained line spectrum is normalized by its dc component to allow interpretation in terms of modulation indices, and binned into 1/3-octave bands in the range from 0.40 to 31.5 Hz. This gives a separate modulation spectrum for each of the audio-frequency octave bands.

As shown in Fig. 2, the envelope spectra for the three different speech materials all have the usual maximum around 3–4 Hz, but show differences in magnitude. Results

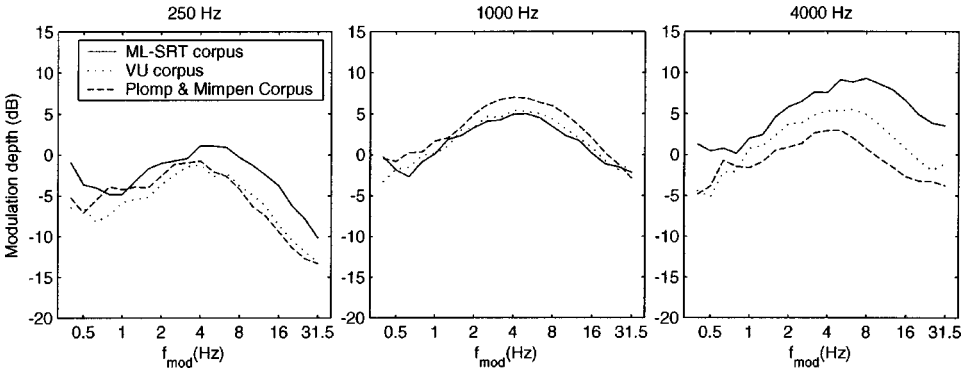


FIG. 2. Averaged envelope spectra (250-, 1000-, and 4000-Hz audio frequency bands) of speech by three different talkers.

appear different for each individual audio-frequency octave band, making it difficult to detect systematic differences due to the speech material.

By inspecting envelope spectra such as Fig. 2, the only (subtle) trend that may be observed is that for the clearer Plomp and Mimpen sentences, the energy in the envelope spectrum appears to be concentrated more around the maximum at 3 Hz. It spreads a smaller fraction of its total energy to higher modulation frequencies.

To investigate whether this is a systematic effect, frequency-integrated versions of the envelope spectra are calculated, averaged across audio frequency, and normalized by dividing through their cumulative maximum (making the value at 31.5 Hz, the highest measured modulation frequency, equal to 1). Figure 3 shows these integrated (or cumulative) spectra for the three different speech materials, integrated from 1 Hz upward.

The tendency in Fig. 3 appears to be that the envelope spectrum of clearer speech shows relatively smaller contributions by the higher modulation frequencies. The modulation frequencies in Fig. 3 not taken into account by the STI model (>12.5 Hz) represent only a small portion of the total energy for the Plomp and Mimpen corpus, but are of greater importance for the ML-SRT and VU material.

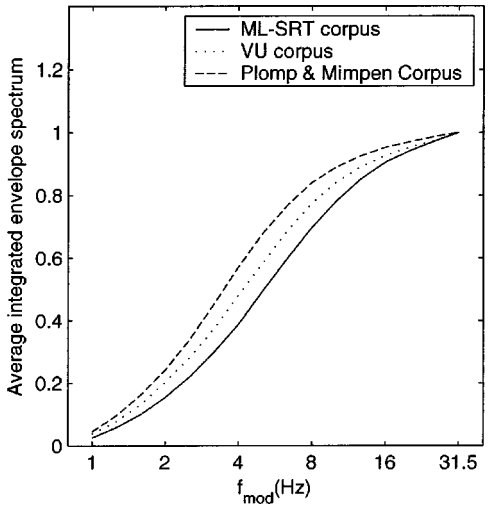


FIG. 3. Integrated (cumulative) envelope spectra of speech by three different talkers. The square of modulation index *m* was integrated from 1 Hz upward, and averaged across the audio frequency octave bands 125–8000 Hz.

C. Adapting the STI method by using a wider modulation frequency range

A straightforward first step in trying to adapt the STI model for more conversational speech would be to extend the modulation frequency range to 31.5 Hz, maintaining equal weight for all modulation frequencies. The modulation frequencies remain separated by 1/3 octave, so the extension to 31.5 Hz increases the number of modulation frequencies from 14 to 18.[1]

Figure 4 is based on the same SRT data as Fig. 1, this time with the modulation frequency range for the STI calculation extended to 31.5 Hz. The ML-SRT data in Fig. 4 show a much closer resemblance to the expected horizontal line than in Fig. 1. The same is true for the VU data, even if some dependence of the STI on the EDT is still observed (the STI at EDT=0.50 differs significantly from the STI without reverberation). Only for the Plomp and Mimpen data, Fig. 1 fits the expected horizontal line better. This confirms the expectations based on the modulation spectra of Fig. 3.

D. Envelope spectra for a larger population of talkers

Given the differences in modulation spectra for the three SRT talkers, the question arises what variations may be expected for a greater population of (arbitrarily selected) talkers.

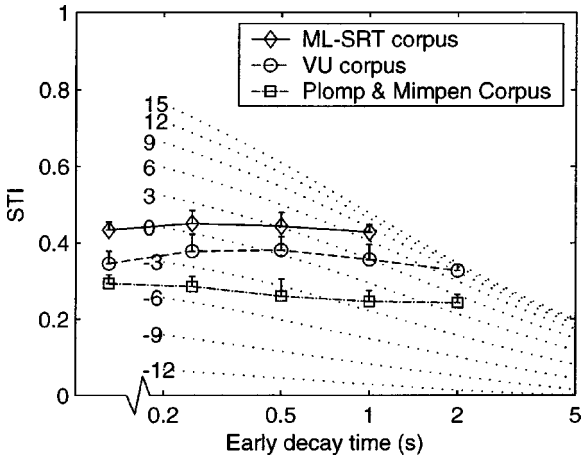


FIG. 4. STI at the SRT results, based on the same data as Fig. 1, but with a wider modulation frequency range (0.63–31.5 Hz). The dotted reference lines (STI vs EDT as a function of SNR) are also based on this extended modulation frequency range.

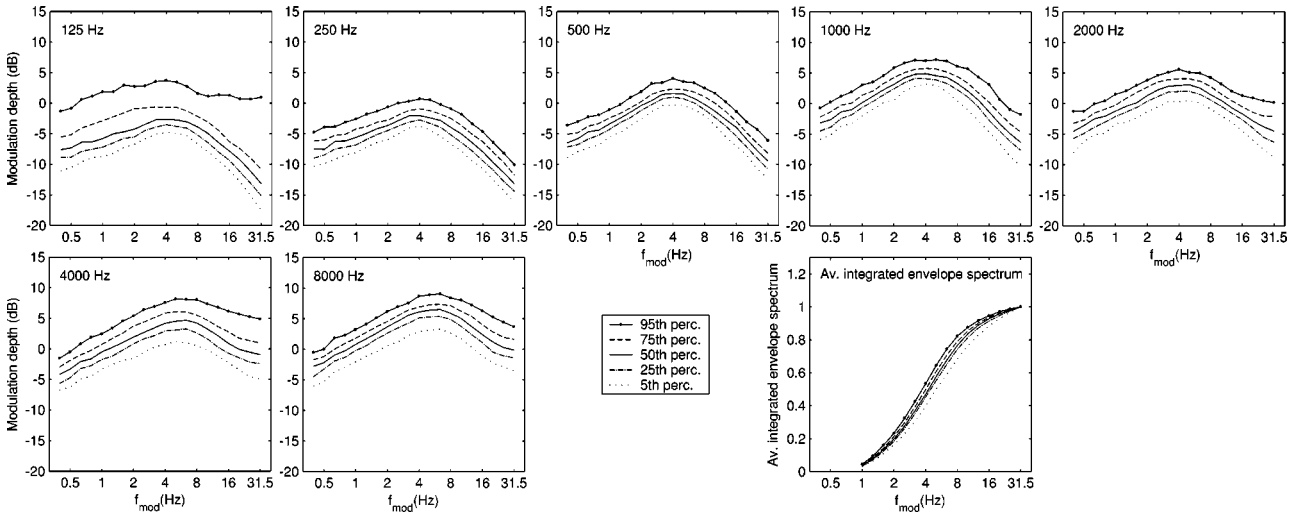


FIG. 5. Envelope spectra (125–8000-Hz audio frequency octave bands) of speech by 134 different talkers. The data are represented by the 5th, 25th, 50th, 75th, and 95th percentile. The corresponding integrated (cumulative) envelope spectra (the square of modulation index *m* integrated from 1 Hz upward and averaged across all audio frequency octave bands) are also given.

Speech material (Dutch newspaper sentences) read aloud by 134 (male and female) native Dutch talkers, taken from the NRC corpus (van Leeuwen and Orr, 2000), was subjected to the same modulation spectrum calculations as the SRT sentences. The NRC corpus consists of high-quality recordings of untrained talkers, screened for impairments, but otherwise randomly selected.

Figure 5 shows that the maximum of the envelope spectrum shifts slightly (from approximately 3 to 4 Hz) for higher audio frequencies. The statistical spread is considerable, especially for the higher frequency bands. The percentile curves were calculated separately for each band. The percentile belonging to a certain talker in a certain band was found to have no predictive value for the percentile this talker would correspond to in other bands; no systematic correlations were found. The bottom right panel of Fig. 5 shows 5th–95th-percentile versions of the integrated envelope spectrum, derived from the individual talker data rather than by integrating the curves shown in the other panels.

The 5th- and 95th-percentile curves in this panel are close to the ML-SRT and Plomp and Mimpen data, respectively, in Fig. 3. This indicates that the Plomp and Mimpen and ML-SRT talkers represent the extremes of the talker population on which Fig. 5 is based.

IV. CONCLUSIONS AND DISCUSSION

Depending on the talker and the adopted speaking style, the standardized STI calculation procedure (IEC, 1998) may give inaccurate predictions of sentence intelligibility in reverberant conditions. Based on the data presented in this paper, we propose to apply a wider range of modulation frequencies (0.63–31.5 Hz instead of 0.63–12.5 Hz) for predicting the intelligibility of conversational speech. For clear speech, the standard modulation frequency range remains more appropriate.

Further fine-tuning of the STI model may be possible through the application of modulation frequency weighting functions that depend on speaking style. For the limited range of variations in speaking style and voice quality addressed in this study, such a refined and complex approach would not be justified. However, more extreme variations in speaking style (including true conversations, where the interaction between the communicators becomes important) may require this more refined approach.

[1]Extension of the range to higher modulation frequencies is one of the ways in which the STI model can be made more sensitive to reverberation. Another possibility would have been to maintain 14 modulation frequencies, but shift the entire range upward. It has been verified that, for the present data, this leads to essentially similar results. However, this would also affect the relation between the STI and intelligibility for conditions affecting the low-frequency end of the envelope spectrum, such as AGC (automatic gain control).

Duquesnoy, A. J. H. M., and Plomp, R. (**1980**). ''Effect of reverberation and noise on the intelligibility of sentences in cases of presbyacusis,'' J. Acoust. Soc. Am. **68**, 537–544.

Houtgast, T., Steeneken, H. J. M., and Plomp, R. (**1980**). ''Predicting speechintelligibility in rooms from the modulation transfer function. I. General room acoustics,'' Acustica **46**, 60–72.

IEC (**1998**). IEC 60268-16 2nd edition, ''Sound system equipment. Part 16: Objective rating of speech intelligibility by speech transmission index'' (International Electrotechnical Commission, Geneva, Switzerland).

Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). ''Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **95**, 1581–1592.

Plomp, R., and Mimpen, A. M. (**1979**). ''Improving the reliability of testing the speech reception threshold for sentences,'' Audiology **18**, 43–52.

Steeneken, H. J. M., and Houtgast, T. (**1980**). ''A physical method for measuring speech transmission quality,'' J. Acoust. Soc. Am. **67**, 318–326.

van Leeuwen, D. A., and Orr, R. (**2000**). ''Speech recognition of non-native speech using native and non-native acoustic models,'' in Proceedings of the RTO workshop MIST, RTO-MP-28 AC/323(IST)TP/4, Neuilly-sur-Seine, France.

van Wijngaarden, S. J., Steeneken, H. J. M., and Houtgast, T. (**2001**). ''Methods and models for quantitative assessment of speech intelligibility in cross-language communication,'' in Proceedings of the RTO Workshop on Multi-lingual Speech and Language Processing, Aalborg, Denmark.

van Wijngaarden, S. J., Steeneken, H. J. M., and Houtgast, T. (**2002**). ''Quantifying the intelligibility of speech in noise for nonnative listeners,'' J. Acoust. Soc. Am. **111**(4), 1906–1916.

Versfeld, N. J., Daalder, J., Festen, J. M., and Houtgast, T. (**2000**). ''Method for the selection of sentence materials for efficient measurement of the speech reception threshold,'' J. Acoust. Soc. Am. **107**, 1671–1684.

The diffraction of sound by an impedance sphere in the vicinity of a ground surface

Kai Ming Li[a] and Wai Keung Lui
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Glenn H Frommer
MTR Corporation Ltd., MTR Tower, Telford Plaza, Kowloon Bay, Hong Kong

(Received 10 January 2003; revised 5 July 2003; accepted 23 September 2003)

The problem of sound diffraction by an absorbing sphere due to a monopole point source was investigated. The theoretical models were extended to consider the case of sound diffraction by an absorbing sphere with a locally reacting boundary or an extended reaction boundary placed above an outdoor ground surface of finite impedance. The separation of variables techniques and appropriate wave field expansions were used to derive the analytical solutions. By adopting an image method, the solutions could be formulated to account for the multiple scattering of sound between the sphere and its image near a flat acoustically hard or an impedance ground. The effect of ground on the reflected sound fields was incorporated in the theoretical model by employing an approximate analytical solution known as the Weyl–van der Pol formula. An approximation solution was suggested to determine the scattering coefficients from a set of linearly coupled complex equations for an absorbing sphere not too close to the ground. The approximate method substantially reduced the computational time for calculating the sound field. Preliminary measurements were conducted to characterize the acoustical properties of an absorbing sphere made of open cell polyurethane foam. Subsequent experiments were carried out to demonstrate the validity of the proposed theoretical models for various source/receiver configurations around the sphere above an acoustically hard ground and an impedance ground. Satisfactory comparative results were obtained between the theoretical predictions and experimental data. It was found that the theoretical predictions derived from the approximate solution agreed well with the results obtained by using the exact solutions. [DOI: 10.1121/1.1628681]

PACS numbers: 43.20.Bi, 43.20.El, 43.28.Js [LCS] Pages: 42–56

I. INTRODUCTION

The problem of acoustic diffraction by an object has attracted the attention of researchers since the late 19th century. Researchers are especially interested in the phenomenon of sound diffraction by spheres because of the relatively simple mathematical models that can be derived. The theoretical models have been used in many applications in fields as diverse as engineering, biology, marine studies, and medicine. In the past decades, there was significant interest in finding the analytical solutions for a plane wave incident on a spherical object, and the studies have recently been extended to include the presence of a hard, soft, or impedance plane boundary.[1–5] More recently, Dassios *et al.*[6] undertook a series of studies on the problems of diffraction by a point source excitation. A full list of their articles and other references are given in Ref. 6. In their work, they propose analytical approximations for most, if not all, possible boundaries at low frequencies. Their studies include the establishment of the transmission conditions for acoustic scattering of the spherical sound field from a point source by the following: a soft or hard sphere, a sphere with an impedance boundary condition on its surface, a penetrable sphere with either a lossy or lossless interior, a soft or hard small sphere with a lossy coating and the corresponding problem with a resistive core surface, and a sphere coated with a penetrable spherical shell. An exact analytical solution for acoustic diffraction by a sphere irradiated by a point source can be found in many textbooks (see, for example Ref. 7).

In the late 1940s, Weiner[8] conducted an experimental investigation of an incident plane wave scattered by a rigid sphere in air. Similar experimental studies of the acoustic scattering by a solid sphere immersed in water were carried out by Hampton and McKinny.[9] On the other hand, Harbold and Steinberg[10] explored the creeping wave theory for sound diffracted by a hard sphere experimentally. Despite these earlier studies, there are relatively few theoretical and experimental studies exploiting the phenomenon of sound diffracted by an absorbing spherical object above an outdoor ground surface of finite impedance. A notable exception is the work of Huang and Gaunaurd[4] on acoustic point source scattering by a spherical elastic shell submerged beneath a free surface in underwater acoustics.

Point-source radiation has a broad application in the fields of underwater acoustics, medical imaging, nondestructive testing, and seismic wave exploration.[6] The current study, however, is motivated by the need to develop an accurate numerical model pertaining to the diffraction of sound by a finite obstacle irradiated by a point source in an urban environment, for instance, the study of sound radiated from

[a]Electronic mail: mmkmli@polyu.edu.hk

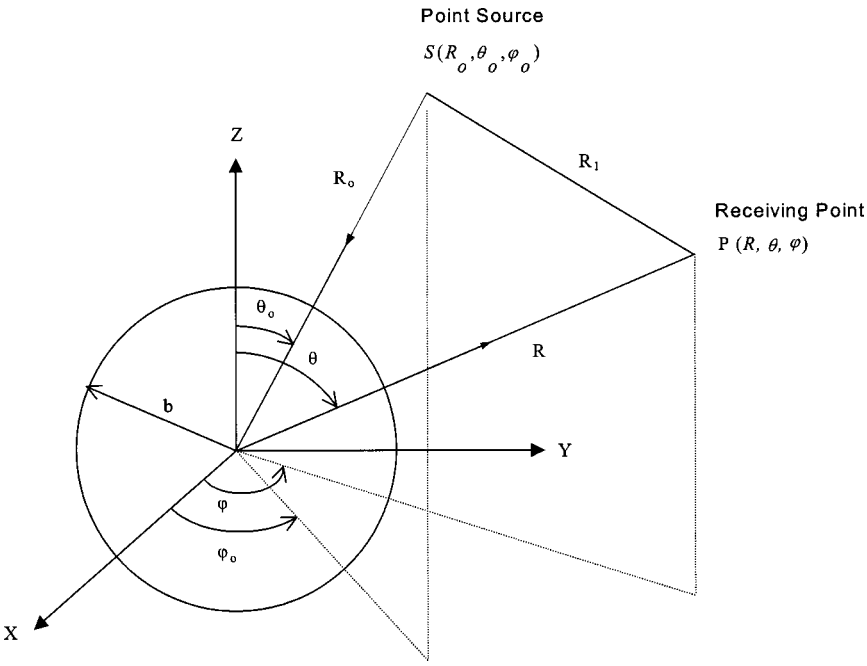


FIG. 1. Illustration of the geometrical configuration of the diffraction problem by a sphere irradiated by a point source.

tires above a porous road pavement.[11] Here, we extend previous analyses to allow the prediction of sound fields by the presence of an absorbing sphere near a flat ground of different boundary conditions. The exact analytical solution for this relatively simple case is needed because it will provide a benchmark for other more complicated and theoretically intractable numerical schemes. The analytical solutions developed in the present work can also be useful in assessing other numerical models used in related fields such as acoustical oceanography and others.

Section II of this paper outlines an analytical solution of acoustic diffraction by an absorbing sphere excited by a point source. We endeavor to explain the physical parameters involved in the mathematical model. In Sec. III the problem is extended to include the presence of a hard ground as well as an impedance ground. The image method[12] and appropriate addition theorem[13–16] as presented by Huang and Gaunaurd[1–4] will be employed to tackle the multi-scattering problem. The effect of ground impedance on the reflected sound waves is incorporated in the numerical model by using an asymptotic solution[17–19] developed for the reflection of a spherical wave from a locally reacting surface. In Sec. IV, experimental measurements of the present study are presented. They are used to validate the theoretical models for various receiver locations around the sphere above the hard and impedance ground.

II. SOUND DIFFRACTION BY A SPHERE IN FREE SPACE

Let a sphere of radius b be placed in an unbounded homogeneous medium. It is insonified by a point source and the total sound field is governed by the Helmholtz equation expressed in terms of a scalar velocity potential, ϕ as

$$\nabla^2\phi+k^2\phi=-\delta(x-x_s), \tag{1}$$

where $k=\omega/c$ is the wave number, ω is the angular frequency of the source, and c is the sound speed in the medium. In Eq. (1), the time-dependent factor $e^{-i\omega t}$ is understood and suppressed throughout the analysis. The problem can be tackled by choosing a spherical coordinate system with the origin selected at the center of the sphere, as shown in Fig. 1. The source point S is located at (R_o,θ_o,φ_o), and the field point P is situated at (R,θ,φ). We are primarily interested in the situation where source and receiver are located at the external positions of the sphere, i.e., R_o, $R\geqslant b$ (the radius of the sphere). The method for obtaining the analytical solution of the sound scattered by an absorbing sphere is given in Ref. 7. We shall not repeat these procedures but, rather, we quote the analytical solution below.

The total sound field due to the presence of a sphere made of either locally reacting material or extended reaction material irradiated by a point source can be determined according to

$$\phi=\sum_{n=0}^{\infty}\sum_{m=0}^{n}[a_{mn}j_n(kR)+b_{mn}h_n^{(1)}(kR)]h_n^{(1)}(kR_o)\times P_n^m(\cos\theta)P_n^m(\cos\theta_o)\cos m(\varphi-\varphi_o)\quad\text{for } R<R_o, \tag{2a}$$

where $j_n(\,)$ and $h_n^{(1)}(\,)$ are, respectively, the spherical Bessel functions of the first and third kinds, and $P_n^m(\,)$ are the associated Legendre functions expressed in terms of hypergeometric functions as

$$P_n^m(z)=\frac{1}{(-m)}\left(\frac{z+1}{z-1}\right)^{(1/2)m}F\left(-n,n+1;1-m;\frac{1}{2}-\frac{1}{2}z\right). \tag{2b}$$

$$F(\alpha,\beta;\gamma;z)=1+\frac{\alpha\beta}{1!\,\gamma}+\frac{\alpha(\alpha+1)\beta(\beta+1)}{2!\,\gamma(\gamma+1)}+\cdots. \tag{2c}$$

The coefficient a_{mn} in Eq. (2a) is determined according to

$$a_{mn}=\frac{i(n-m)!(2n+1)(2-\delta_{m0})k}{4\pi(n+m)!}, \tag{3}$$

and δ_{m0} is a Kronecker delta function that vanishes if $m \neq 0$. The roles of R and R_o in Eq. (2) can be interchanged when $R > R_o$ because of the reciprocity condition between source and receiver. The unknown scattering coefficients b_{mn} are to be determined by imposing proper boundary conditions on the surface of the sphere. For an extended reaction sphere, the scattering coefficients are given by in terms of a_{mn} as

$$b_{mn} = -\frac{j_n'(kb) - \zeta\eta(j_n'(k_1b)/j_n(k_1b))j_n(kb)}{h_n^{(1)\prime}(kb) - \zeta\eta(j_n'(k_1b)/j_n(k_1b))h_n^{(1)}(kb)} a_{mn}, \tag{4}$$

where ζ is the density ratio and η is the index of refraction of the spherical surface. The primes in the spherical Bessel functions of Eq. (4) denote the derivatives with respect to their arguments.

We find it convenient to introduce a term called effective acoustic admittance β_e, which is related to the specific normalized acoustic admittance β_s of the sphere. They are defined by

$$\beta_s = \zeta\eta \tag{6}$$

and

$$\beta_e = \beta_s \frac{j_n'(k_1b)}{j_n(k_1b)}. \tag{7}$$

It is remarkable that when the sound speed in the sphere is much smaller than the sound speed in air, the argument k_1b will be very large and the ratio $j_n'(k_1b)/j_n(k_1b)$ can be approximated by $-i$. This approximation can be shown by invoking the asymptotic expansion of the spherical Bessel function $j_n(Z)$, the tangent function in terms of real and imaginary parts, and by taking the limiting values of hyperbolic functions. With such an approximation, the scattering coefficients can be reduced:

$$b_{mn} = -\frac{j_n'(kb) + i\beta_s j_n(kb)}{h_n^{(1)\prime}(kb) + i\beta_s h_n^{(1)}(kb)} a_{mn}. \tag{8}$$

This type of boundary is called a locally reacting boundary because the sound waves at any angle of incidence will be strongly refracted along the radius and thus travel normal to the spherical surface. Indeed, if we apply an impedance boundary condition at the surface of the sphere, i.e., $\partial\phi/\partial R + ik\beta_s\phi = 0$ at $R = b$, then we can deduce the same scattering coefficient as shown in Eq. (8) above. This implies that a local reaction model would be a good approximation for the extended reaction model when the propagation constant of the porous materials is sufficiently large. We shall elaborate on this argument in the following paragraphs by providing some numerical examples. However, we point out a limiting case that if $kb \gg 1$, i.e., at sufficient frequency or a sphere of a small radius, then the numerical results can be reduced to the classical Rayleigh scattering approximation.

The validity of computing sound fields by approximating a sphere of an extended reaction boundary as a locally reacting one when the propagation constant of the sphere is large can be assessed conveniently by employing a singlc-parameter model.[20] The model, which is also known as the Delaney and Bazley model, describes the propagation constant and normalized surface impedance of a fibrous porous material in terms of a ratio of the source frequency f to the effective flow resistivity σ_e of the material. It is chosen because of its simplicity as only a single parameter is needed to compute the relevant parameters for either a locally reacting boundary or an extended reaction one. We note that there are other more sophisticated models involving two or more parameters, but the Delaney and Bazley model is sufficient in this section. With the numerical simulations, we can estimate the lower limit of the effective flow resistivity at which an extended reaction model can be adequately approximated by a locally reacting model.

Figure 2 shows the numerical comparisons for the sound diffracted by a sphere with a radius of 0.1 m in the absence of other reflecting surfaces. In the plots, the dotted lines represent the predicted sound fields in the vicinity of an extended reaction sphere and the solid lines are the corresponding predictions for a locally reacting sphere. In Figs. 2(d) and (g), we also show the respective plots for an acoustically hard sphere of the same radius. In these plots, the sound fields are determined by using Eq. (2a) with the scattering coefficient determined either by Eq. (4) or Eq. (8) for an extended reaction material or locally reacting material, respectively. For a consistent presentation, the relative sound pressure level is used throughout the current study. It is defined as the ratio of the total sound field relative to the reference free-field level measured at 1 m from the source. Two representative receiver locations are chosen in the plots: one located in the illuminated zone with $R = 0.15$ m, $\theta = 90°$, $\varphi = 30°$, [see Figs. 2(a)–(d)] and the other located in the shadow zone with $R = 0.15$ m, $\theta = 90°$, $\varphi = 150°$ [see Figs. 2(e)–(h)]. Different values of the effective flow resistivity σ_e of 5, 25, 50, and 100 kPa s m^{-2} are selected in our numerical simulations in order to estimate an approximate lower limit of the effective flow resistivity at which an extended reaction model can be approximated by a local reaction model. For the material used with $\sigma_e \geqslant 50$ kPa s m^{-2}, it is found that the predicted sound fields from both types of surfaces are in close agreement with each other (see Fig. 2). These comparisons suggest that a locally reacting model is an adequate approximation if the effective flow resistivity of the absorbing sphere is greater than 50 kPa s m^{-2}.

III. SOUND DIFFRACTION ABOVE AN OUTDOOR GROUND

The sound wave propagation mechanisms, and hence the sound field above an outdoor ground, are inevitably affected by the acoustical properties of the ground surface, principally the acoustic impedance of the ground.[21] The acoustical characteristics of the ground determine the phase and amplitude of reflected waves, which interfere with direct waves to produce an intricate sound field. A handy model for determining the acoustical properties of a ground is therefore desired in addition to a scattering model for predicting a sound field induced by a scatterer above an outdoor ground surface. The current scattering problem is formulated by considering a sphere of radius b at a distance $d/2$ above a flat ground. The geometrical configuration of the scattering problem is shown in Fig. 3. The spherical coordinate system is employed with

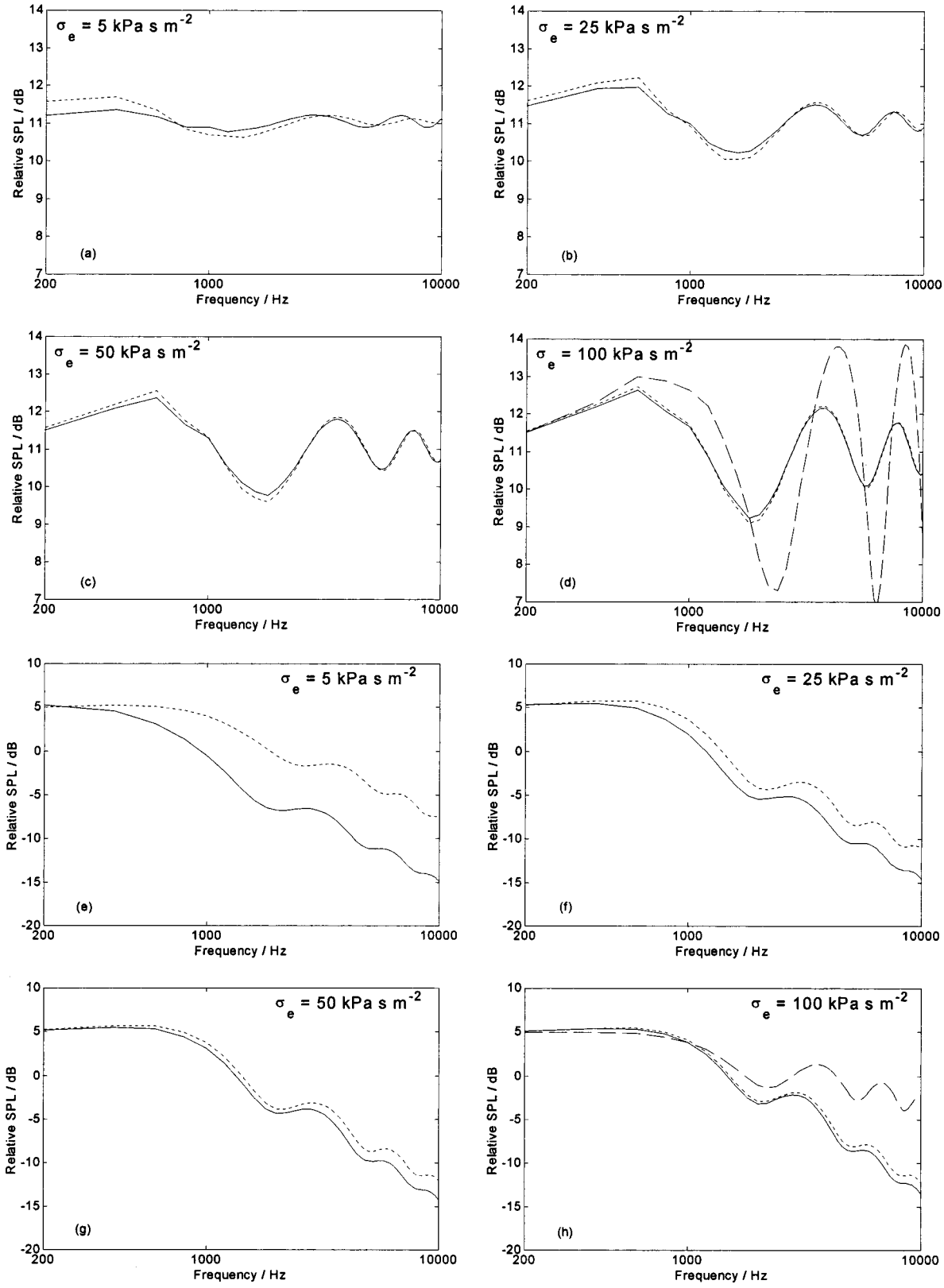


FIG. 2. Numerical comparisons for the acoustic diffraction by a sphere with a radius of 0.1 m using the Delaney and Bazley locally reacting model and an extended reaction model. Different values of effective flow resistivity σ_e are used for comparison. Solid line: locally reacting model. Dotted line: extended reaction model. Source: $R_o = 0.4$ m, $\theta_o = 90°$, $\varphi_o = 0°$. Receiver locations: (a)–(d), $R = 0.15$ m, $\theta = 90°$, $\varphi = 30°$; (e)–(h), $R = 0.15$ m, $\theta = 90°$, $\varphi = 150°$. The dashed lines in (d) and (h) represent the relative SPL for an acoustically hard sphere with $\sigma_e = 30\,000$ kPa s m^{-2}.

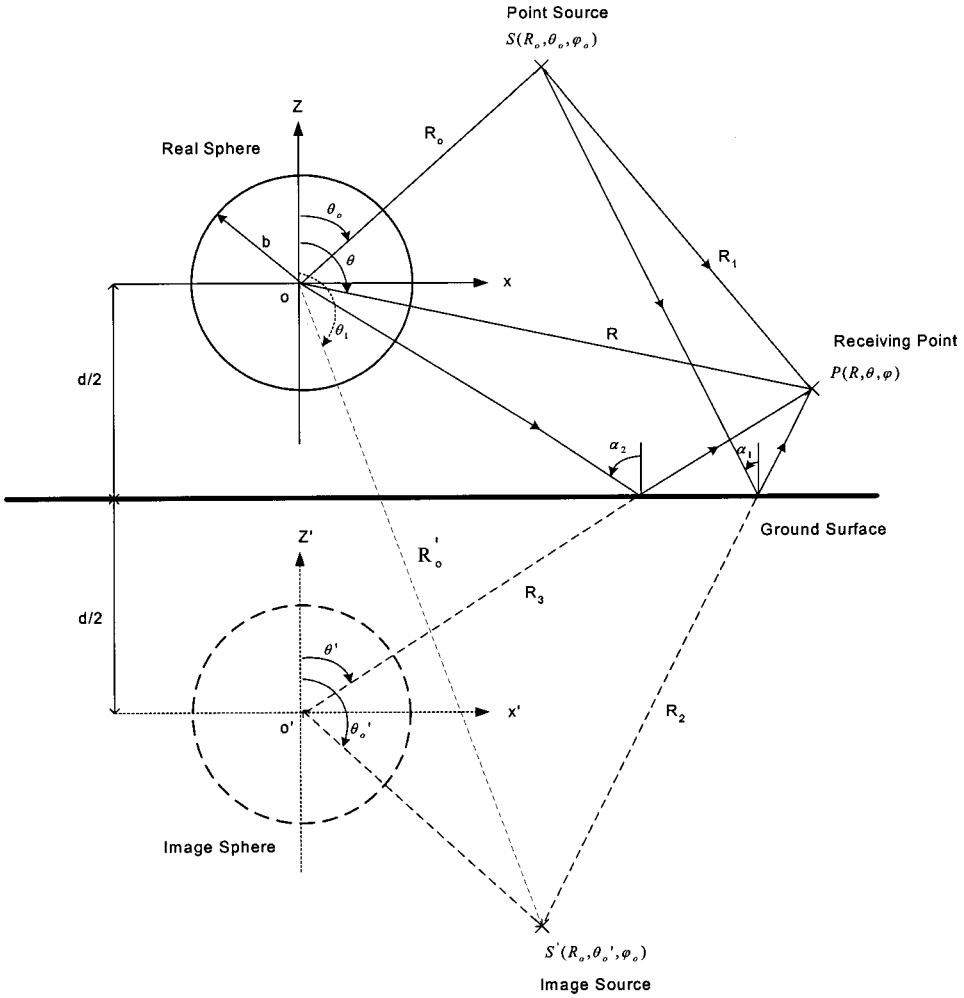


FIG. 3. Illustration of the geometrical configuration of the acoustic diffraction problem by a sphere above a flat ground irradiated by a point source.

the center of the sphere taken as the origin. The point source is positioned at a finite distance from the ground and irradiating on the sphere. The sound field anywhere is inherently contributed by the direct radiation of the source and the exterior radiation (diffraction) from the sphere, as well as by reflections from the outdoor ground surface.

A. Diffraction of sound by a sphere above a hard ground

The sound field produced due to a sphere irradiated by a point source in a free field can be computed by using Eq. (2a). The contributions due to the reflected sound waves can be readily found by applying the image method in accordance with the theory of geometrical acoustics. In this case, the problem is tackled by expressing the sound field in four components: the direct source, scattered waves from the sphere, waves disseminating from the image source, and the image sphere (see Fig. 3). The sound field contributed by an image source in a hard ground can be expanded in an infinite series of associated Legendre functions and spherical Bessel functions as

$$\phi_r^i = \sum_{n=0}^{\infty} \sum_{m=0}^{n} a_{mn} h_n^{(1)}(kR_0') j_n(kR) P_n^m(\cos\theta_1) \times P_n^m(\cos\theta)\cos m(\varphi-\varphi_o), \tag{9}$$

where R_o' is the distance measured from the center of the real sphere to the image source, and the constant a_{mn} is given by Eq. (3). By considering $\theta_o' = \pi - \theta_o$, and using the identity of Legendre functions $P_n^m(\cos\theta_o') = (-1)^{m+n} P_n^m(\cos\theta_o)$, the contribution due to the scattering sound field from the image sphere in a hard ground can be represented in an analogous form

$$\phi_r^s = \sum_{q=0}^{\infty} \sum_{m=0}^{q} (-1)^{m+q} b_{mq} h_q^{(1)}(kR_o) \times h_q^{(1)}(kR_3) P_q^m(\cos\theta_o) P_q^m(\cos\theta')\cos m(\varphi-\varphi_o), \tag{10}$$

where the unknown coefficients b_{mq} have to be determined from the boundary condition on the spherical surface. By the principle of superposition, we can express the total sound field as the sum of direct source, scattered waves from the real sphere, image source, and scattered waves from the image sphere,

$$\phi^T = \phi^i + \phi^s + \phi_r^i + \phi_r^s. \tag{11}$$

Since the sound field contributed from an image sphere is still expressed in terms of spherical polar coordinates where the origin is at the center of the image sphere "O'," we need to express this solution in a real spherical coordinate system with the origin at the center of the "real" sphere "O." The

analysis can be simplified somewhat by employing the translational addition theorem[13,14] for the spherical wave functions as presented by Huang and Gaunaurd.[1–4] For the case of vertical translation, we have a simplified formulation for the addition theorem:[15,16]

$$h_q^{(1)}(kR_3)P_q^m(\cos\theta')=\sum_{n=m}^{\infty} A_{mn}^{mq}(kd)j_n(kR)P_n^m(\cos\theta), \tag{12a}$$

where the translation coefficients

$$A_{mn}^{mq}(kd)=(-1)^m i^{n-q}(2n+1)\sum_p i^p a(m,q,-m,n,p) \times\begin{bmatrix} h_p^{(1)}(kd) \\ j_p(kd)\end{bmatrix},\quad \begin{bmatrix} R\leqslant d \\ R>d\end{bmatrix} \tag{12b}$$

with $p=q+n,q+n-2,...,|q-n|$ and the Gaunt coefficients given by

$$a(m,q,-m,n,p)=(2p+1)\times\left[\frac{(q+m)!(n-m)!}{(q-m)!(n+m)!}\right]^{1/2}\begin{bmatrix} q & n & p \\ 0 & 0 & 0\end{bmatrix}\times\begin{bmatrix} q & n & p \\ m & -m & 0\end{bmatrix}. \tag{12c}$$

The Gaunt coefficients are expressed as products of the Wigner $3-j$ symbols [the last two terms of Eq. (12c)]. The Wigner $3-j$ symbol, which is associated with the coupling of two angular momentum eigenvectors, is defined in the Van der Waerden form[22] by

$$\begin{bmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3\end{bmatrix}=(-1)^{j_1-j_2-m_1-m_2}\Delta(j_1\quad j_2\quad j_3) \times\sqrt{(j_1+m_1)!(j_1-m_1)!(j_2+m_2)!(j_2-m_2)!(j_3+m_1+m_2)!(j_3-m_1-m_2)!} \times\sum_n \frac{(-1)^n}{n!(j_1+j_2-j_3-n)!(j_1-m_1-n)!(j_2+m_2-n)!(j_3-j_2+m_1+n)!(j_3-j_1+m_2+n)!}, \tag{13a}$$

where the ''triangle relation'' is defined by

$$\Delta(j_1\quad j_2\quad j_3)=\sqrt{\frac{(j_1+j_2-j_3)!(j_1-j_2+j_3)!(-j_1+j_2+j_3)!}{(j_1+j_2+j_3+1)!}}. \tag{13b}$$

The summation index n assumes all integer values for which the factorial arguments are not negative. The Gaunt coefficients can be either directly calculated using Eq. (12c) or evaluated by the recursion relation developed by Bruning and Lo[15,16] (see the Appendix). However, use of the recursion relation is preferred because its use can substantially reduce computational time.

The total sound field due to a sphere irradiated by a point source above a hard ground can now be expressed in a real spherical coordinate system which consists of four components: the direct waves from the source and sphere, and the scattered waves from the image source and image sphere. Mathematically, the sound field can be written in a form as

$$\begin{aligned}\phi^T=&\sum_{n=0}^{\infty}\sum_{m=0}^{n} a_{mn}h_n^{(1)}(kR_o)j_n(kR)P_n^m(\cos\theta)P_n^m(\cos\theta_o)\cos m(\varphi-\varphi_o)\\ &+\sum_{n=0}^{\infty}\sum_{m=0}^{n} a_{mn}h_n^{(1)}(kR_o')j_n(kR)P_n^m(\cos\theta)P_n^m(\cos\theta_o)\cos m(\varphi-\varphi_o)\\ &+\sum_{n=0}^{\infty}\sum_{m=0}^{n} b_{mn}h_n^{(1)}(kR_o)h_n^{(1)}(kR)P_n^m(\cos\theta)P_n^m(\cos\theta_o)\cos m(\varphi-\varphi_o)\\ &+\sum_{q=0}^{\infty}\sum_{m=0}^{q}(-1)^{m+q}b_{mq}h_q^{(1)}(kR_o)P_q^m(\cos\theta_o)\sum_{n=m}^{\infty}A_{mn}^{mq}(kd)j_n(kR)P_n^m(\cos\theta)\cos m(\varphi-\varphi_o).\end{aligned} \tag{14}$$

For an absorbing sphere with a locally reacting boundary, the unknown scattering coefficients b_{mn} can be determined by the following equation,

$$\frac{\partial\phi^T}{\partial R}+ik\beta_s\phi^T=0,\quad \text{at } R=b, \tag{15}$$

where β_s is the specific normalized acoustic impedance of the sphere as defined in Eq. (6). After equating the coefficients of $P_n^m(\cos\theta)\cos m(\varphi-\varphi_o)$ and some tedious manipulations, a set of infinite and coupled linear complex equations can be formed for solving b_{mn}. The set of complex equations can be written in the following form:

$$b_{mn}h_n^{(1)}(kR_o)P_n^m(\cos\theta_o)+T_n\sum_{q=0}^{\infty}(-1)^{m+q}b_{mq}h_q^{(1)}(kR_o)\times P_q^m(\cos\theta_o)A_{mn}^{mq}(kd)$$
$$=-a_{mn}T_n[h_n^{(1)}(kR_o)P_n^m(\cos\theta_o)+h_n^{(1)}(kR_o')P_n^m(\cos\theta_1)]. \quad (16)$$

It can be expressed in a more compact form by using matrix notations as

$$\mathbf{XB}=\mathbf{C}, \quad (17)$$

where the diagonal elements of $\mathbf{X}$ are

$$X_{nn}=h_n^{(1)}(kR_o)P_n^m(\cos\theta_o)+T_n(-1)^{m+n}\times h_n^{(1)}(kR_o)P_n^m(\cos\theta_o)A_{mn}^{mn}(kd), \quad (18a)$$

the off-diagonal elements of $\mathbf{X}$ are

$$X_{nq}=T_n(-1)^{m+q}h_q^{(1)}(kR_o)P_q^m(\cos\theta_o)A_{mn}^{mq}(kd), \quad (18b)$$

and the elements of the vector $\mathbf{C}$ are

$$C_n=-a_{mn}T_n[h_n^{(1)}(kR_o)P_n^m(\cos\theta_o)+h_n^{(1)}(kR_o')P_n^m(\cos\theta_1)], \quad (18c)$$

with

$$T_n=\frac{j_n'(kb)+i\beta_s j_n(kb)}{h_n^{(1)\prime}(kb)+i\beta_s h_n^{(1)}(kb)}. \quad (18d)$$

The system of complex equations in Eq. (16) can be truncated to an order of N depending on the degree of accuracy required. The order N is the maximum number of terms required in the summation series. Thus, the matrix $\mathbf{X}$ is a $(N+1-m)\times(N+1-m)$ complex matrix, and $\mathbf{B}$ and $\mathbf{C}$ are complex vectors of dimension $(N+1-m)$ for each m where m ranges from 0 to N.

For an absorbing sphere of an extended reaction boundary, it is requisite that the following conditions on the sphere interface be satisfied:

$$\phi^T=\phi^{in}, \quad (19a)$$

$$\frac{\partial\phi^T}{\partial R}=\zeta\frac{\partial\phi^{in}}{\partial R}. \quad (19b)$$

Similarly, a set of infinite and coupled linear complex equations as given in Eq. (16) can be formed for solving b_{mn}. The equations can also be conveniently put in matrix notations in the forms of Eqs. (17) and (18). However, in this case T_n is given by

$$T_n=\frac{j_n'(kb)-i\beta_e j_n(kb)}{h_n^{(1)\prime}(kb)-i\beta_e h_n^{(1)}(kb)}, \quad (20)$$

where β_e is the effective acoustic admittance as defined in Eq. (7).

It is obvious from Eq. (14) that the scattering effect due to the sphere after reflection from the ground is directly related to the translation coefficients $A_{mn}^{mq}(kd)$ which in turn are proportional to $h_p^{(1)}(kd)$ for $R\leqslant d$ or $j_p(kd)$ for $R>d$ as shown in Eq. (12b). The scattering effect from the image sphere depends on both the wave number and the separation distance d between the real sphere and image sphere. Consider a limiting case of $kd\gg 1$. The spherical Bessel functions can be simplified by means of asymptotic expansions to give

$$h_n^{(1)}(kd)\approx i^{-(n+1)}\frac{e^{ikd}}{kd}, \quad (21a)$$

$$j_n(kd)\approx\frac{\cos[kd-(n+1)\pi/2]}{kd}. \quad (21b)$$

Since the spherical Bessel functions in Eqs. (21a) and (21b) decrease with the increase of the parameter kd, and their magnitudes are less than one for $kd\gg 1$, it is therefore expected that the contribution of the ground reflected scattered wave to the total sound field becomes less significant when the sphere is at a larger separation from the ground. Consider an example of an urban scenario: an enclosed section of an elevated road or a railway viaduct located in close proximity to residential buildings. These flyovers and viaducts are usually built at a height of over 10 m above porous road pavements. In this case, the viaduct or the flyover is an ''obstacle'' to the traffic noise having dominant frequencies at about 500 to 2000 Hz. Obviously, the condition of $kd\gg 1$ can be satisfied easily in this case.

Assuming $kd\gg 1$, we can make the following approximation. The off-diagonal elements X_{nq} of Eq. (18b) are negligible compared to the diagonal element X_{nn} in the matrix $\mathbf{X}$ in Eq. (18a). As a result, Eq. (16) can be decoupled, leading to a direct solution for the scattering coefficient b_{mn}, as follows:

$$b_{mn}=\frac{C_n}{X_{nn}}=\frac{-a_{mn}T_n[h_n^{(1)}(kR_o)P_n^m(\cos\theta_o)+h_n^{(1)}(kR_o')P_n^m(\cos\theta_1)]}{h_n^{(1)}(kR_o)P_n^m(\cos\theta_o)+T_n(-1)^{m+n}h_n^{(1)}(kR_o)P_n^m(\cos\theta_o)A_{mn}^{mn}(kd)}. \quad (22)$$

When the sphere is far off the surface of the hard ground, the separation distance d and the distance R_o' from the center of the real sphere to the image source will be very large such that both the spherical Hankel function $h_n^{(1)}(kR_o')$ and the translation coefficient $A_{mn}^{mn}(kd)$ tend to zero. Thus, Eq. (14) can be reduced to Eq. (2a) and Eq. (22) can be simplified to either Eq. (4) or Eq. (8) depending on the boundary condition of the absorbing sphere. These equations are the known expressions for an absorbing sphere in an unbounded medium. In the next section, we shall turn our attention to con-

sider a more complicated situation where the outdoor ground surface has finite impedance.

B. Diffraction of sound by a sphere above an impedance ground

When a sound wave propagates above an impedance ground, the acoustical properties of the ground surface are crucial in determining the sound field produced owing to the combination of the direct and boundary-reflected sound waves. The effect of the ground impedance on the radiation from a point source can be closely modeled by an approximate analytical solution proposed in Refs. 17–19 for the reflection of a spherical wave from a locally reacting plane surface. The sound field in this situation can be expressed in the classical Weyl–van der Pol formula

$$\phi_T^i = \frac{e^{ikR_1}}{4\pi R_1} + Q_1 \frac{e^{ikR_2}}{4\pi R_2}, \tag{23}$$

where Q_1 is the spherical wave reflection coefficient for the point source reflection on the impedance ground. Knowing the distance R_2 between the image source and field point, the angle of incidence of the reflected wave α_1 (see Fig. 3), and the specific normalized admittance of the reflecting ground β_g, the spherical wave reflection coefficient can be determined by

$$Q_1 = V_1 + (1 - V_1) F_1(w_1), \tag{24a}$$

where V_1 is the plane wave reflection coefficient given by

$$V_1 = \frac{\cos\alpha_1 - \beta_g}{\cos\alpha_1 + \beta_g}. \tag{24b}$$

The function $F_1(w_1)$ in Eq. (24a) is known as the boundary loss factor determined by

$$F_1(w_1) = 1 + i\sqrt{\pi} w_1 e^{-w_1^2} \operatorname{erfc}(-i w_1), \tag{24c}$$

where erfc( ) is the complementary error function and the parameter w_1, the so-called numerical distance, is defined as

$$w_1 = +\sqrt{\frac{1}{2} ikR_2}(\cos\alpha_1 + \beta_g). \tag{24d}$$

Since the scattered waves from the sphere are spherically spreading waves, the approximate formulation presented in Refs. 17–19 can be adopted to solve the multiscattering problem in the presence of an impedance ground. This approach was previously used by Hasheminejad[23] to find the modal acoustic force on a spherical radiator near a locally reacting boundary. Following similar arguments in Secs. II and III, a heuristic formula for the multi-scattering sound field due to the presence of an impedance ground can be estimated by

$$\phi_T^s = \phi^s + Q_2 \phi_r^s, \tag{25}$$

where ϕ^s is the scattered sound field from a real sphere in free space, see Ref. 7, and ϕ_r^s is given by Eq. (10). The spherical wave reflection coefficient Q_2 for a scattered wave reflection on an impedance ground can be obtained in a similar formulation as Eq. (24a) with the appropriate plane wave reflection coefficient and boundary loss factor. The total sound field above an impedance ground, which is composed

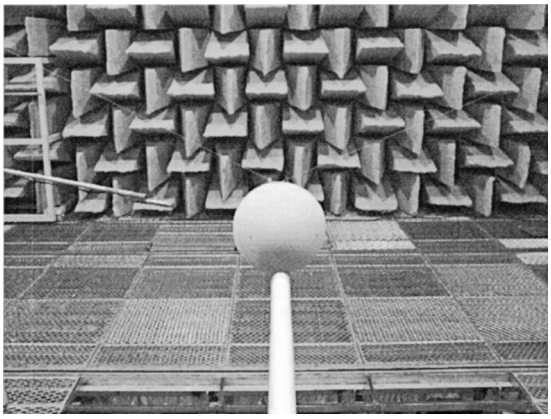


FIG. 4. The experimental setup in an anechoic chamber for the measurement of sound diffraction by an absorbing sphere in a free field. The sphere is bolted with small hooks and hung in the anechoic chamber by suspending fine threads fixed in position by means of poles.

of four terms, can be expressed exactly in similar form as Eq. (14) except for the addition of Q_1 and Q_2 as multipliers for the second and fourth terms, respectively. We can now get another set of coupled linear complex equations for solving the scattering coefficients b_{mn} which are valid for an absorbing sphere of either a locally reacting boundary or an extended reaction boundary,

$$\begin{aligned} & b_{mn} h_n^{(1)}(kR_o) P_n^m(\cos\theta_o) + Q_2 T_n \sum_{q=0}^{\infty} (-1)^{m+q} \\ & \times b_{mq} h_q^{(1)}(kR_o) P_q^m(\cos\theta_o) A_{mn}^{mq}(kd) \\ & = -a_{mn} T_n [h_n^{(1)}(kR_o) P_n^m(\cos\theta_o) \\ & \quad + Q_1 h_n^{(1)}(kR_o') P_n^m(\cos\theta_1)]. \end{aligned} \tag{26}$$

The complex equations can be expediently put in a matrix form similar to Eq. (17). The diagonal elements of **X** in this case are

$$\begin{aligned} X_{nn} = & h_n^{(1)}(kR_o) P_n^m(\cos\theta_o) + Q_2 T_n (-1)^{m+n} \\ & \times h_n^{(1)}(kR_o) P_n^m(\cos\theta_o) A_{mn}^{mn}(kd), \end{aligned} \tag{27a}$$

the off-diagonal elements of **X** are

$$X_{nq} = Q_2 T_n (-1)^{m+q} h_q^{(1)}(kR_o) P_q^m(\cos\theta_o) A_{mn}^{mq}(kd), \tag{27b}$$

and the elements of the vector **C** are

$$\begin{aligned} C_n = & -a_{mn} T_n [h_n^{(1)}(kR_o) P_n^m(\cos\theta_o) \\ & + Q_1 h_n^{(1)}(kR_o') P_n^m(\cos\theta_1)], \end{aligned} \tag{27c}$$

where T_n is determined either by Eq. (18d) for an absorbing sphere with a locally reacting boundary or by Eq. (20) for an extended reaction boundary. In a similar fashion as shown in the last section, the system of complex equations in Eq. (26) can be truncated to an order of N, that is, the number of sums from n or $q=0$ to N, depending on the degree of accuracy required. The complex matrix **X** has an order of $(N+1-m)\times(N+1-m)$ and the complex vectors **B** and **C** have dimensions $(N+1-m)$ for each m where m ranges from 0 to N.

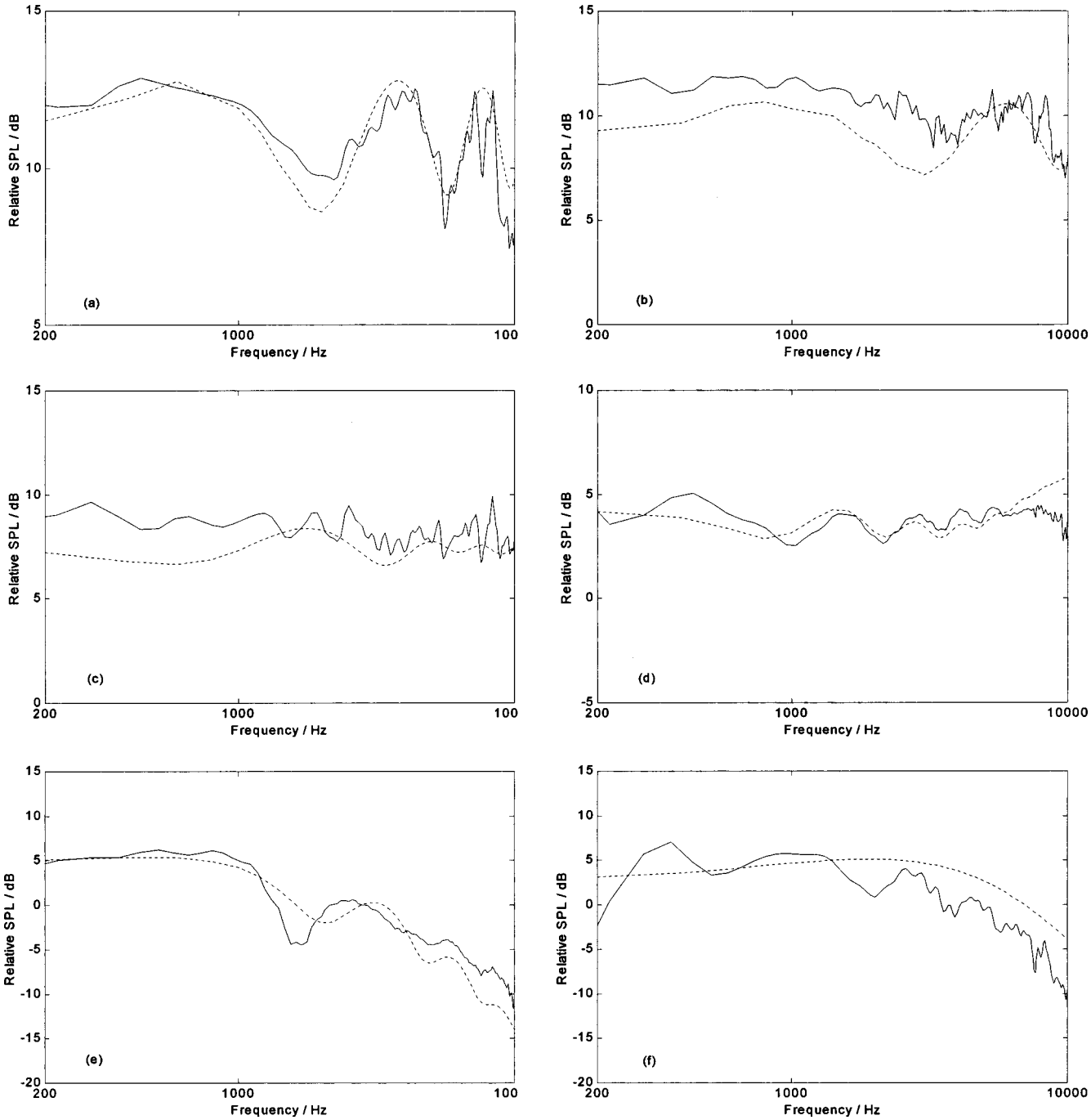


FIG. 5. Experimental results and theoretical predictions for acoustic diffraction by a locally reacting sphere with a radius of 0.1 m in a free space. Source: $R_o=0.4$ m, $\theta_o=90°$, $\varphi_o=0°$. Receiver: (a) $R=0.15$ m, $\theta=90°$, $\varphi=30°$; (b) $R=0.15$ m, $\theta=90°$, $\varphi=60°$; (c) $R=0.15$ m, $\theta=90°$, $\varphi=90°$; (d) $R=0.3$ m, $\theta=90°$, $\varphi=120°$; (e) $R=0.15$ m, $\theta=90°$, $\varphi=150°$; and (f) $R=0.3$ m, $\theta=90°$, $\varphi=180°$. (Dotted line: the theoretical predictions using the two-parameter model with the best-fit parameters are found to be $\sigma_e=100$ kPa s m^{-2} and $\alpha_e=10$ m^{-1}; solid line: experimental result.)

Using the same arguments in the previous section, an approximation can be made. The off-diagonal elements X_{nq} of Eq. (27b) are negligibly small compared with the diagonal element X_{nn} of Eq. (27a) in the matrix $\mathbf{X}$ provided that $kd \gg 1$. The linear complex equation Eq. (26) for b_{mn} can then be decoupled and the scattering coefficient can be solved directly as

$$b_{mn}=\frac{C_n}{X_{nn}}=\frac{-a_{mn}T_n[h_n^{(1)}(kR_o)P_n^m(\cos\theta_o)+Q_1h_n^{(1)}(kR_o')P_n^m(\cos\theta_1)]}{h_n^{(1)}(kR_o)P_n^m(\cos\theta_o)+Q_2T_n(-1)^{m+n}h_n^{(1)}(kR_o)P_n^m(\cos\theta_o)A_{mn}^{mn}(kd)}. \tag{28}$$

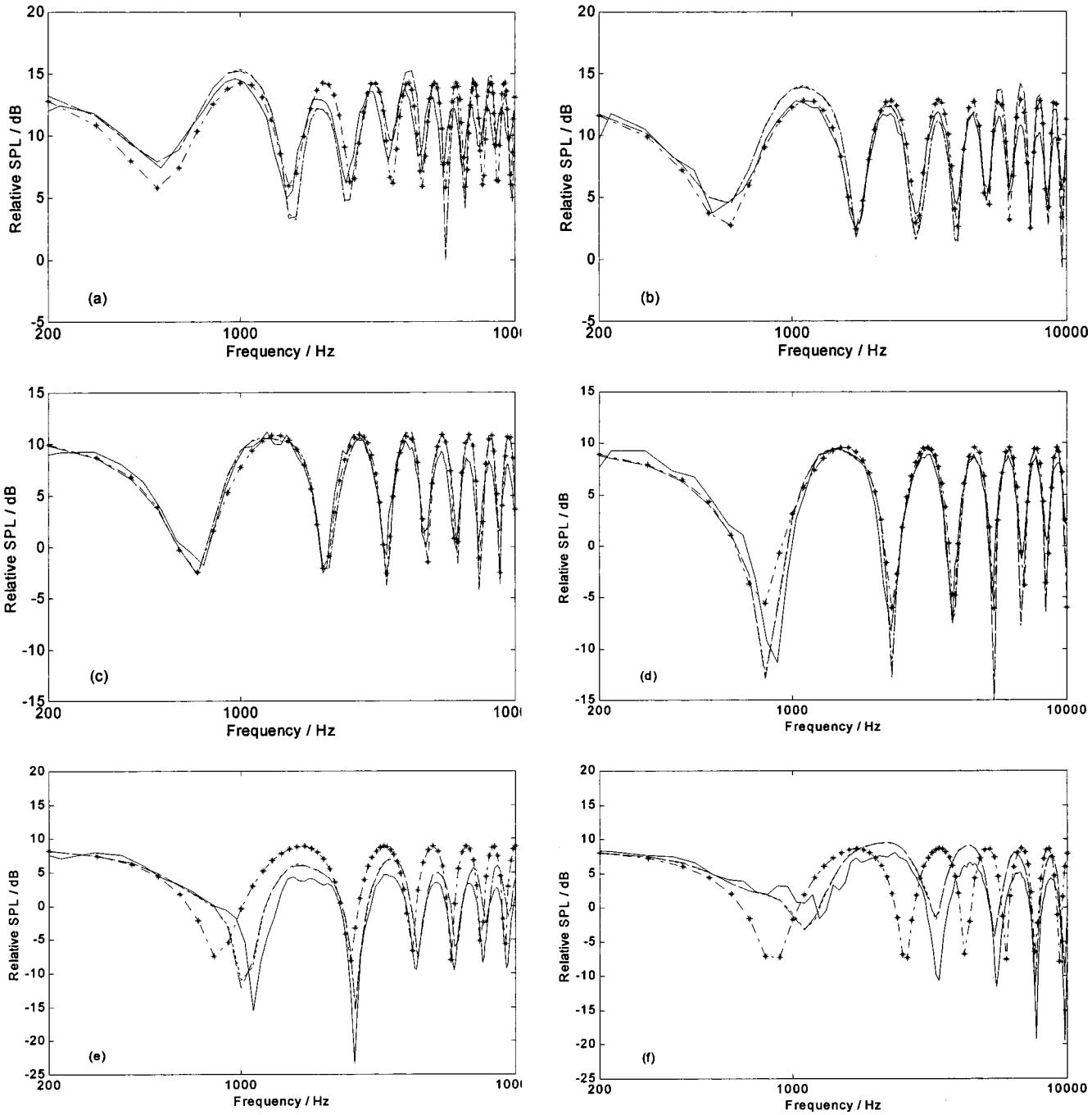


FIG. 6. Experimental results and theoretical predictions for acoustic diffraction by a locally reacting sphere with a radius of 0.1 m above a hard ground. The center of the sphere is 0.275 m above the hard ground. Source: $R_o=0.4$ m, $\theta_o=90°$, $\varphi_o=0°$. Receiver: (a) $R=0.15$ m, $\theta=90°$, $\varphi=30°$; (b) $R=0.15$ m, $\theta=90°$, $\varphi=60°$; (c) $R=0.25$ m, $\theta=90°$, $\varphi=90°$; (d) $R=0.25$ m, $\theta=90°$, $\varphi=120°$; (e) $R=0.25$ m, $\theta=90°$, $\varphi=150°$; and (f) $R=0.25$ m, $\theta=90°$, $\varphi=180°$. (Dotted line: theoretical prediction—approximation; dashed line: theoretical prediction—exact solution; solid line: experimental result; dash-dotted line with legend: predicted sound field without presence of sphere.)

Considering the limiting case when the sphere is very remote from the impedance ground, the contribution from the image source and image sphere to the total sound field are negligible such that Eq. (28) can be reduced to Eq. (4) or Eq. (8), respectively, depending on the boundary condition of the absorbing sphere. It is reassuring to start from a more general situation that leads to an expression for the scattering of sound by an absorbing sphere in an unbounded medium.

In the following sections, we shall present the calculated results that will be used to compare with experimental data. The comparisons provide a validation of the proposed theoretical formulations for the scattering of sound by an absorbing spherical obstacle in the context of outdoor sound propagation.

IV. EXPERIMENTAL AND NUMERICAL RESULTS

A. Instrumentation

A series of indoor experiments were conducted in an anechoic chamber 6 m×6 m×4 m (high) in size for the

validation of the theoretical models introduced in Secs. II and III. The anechoic chamber used in the experiments has been designed with a cut-off frequency of 75 Hz. The point source was produced by means of a tube of 1 m in length and 3 cm in internal diameter. It was driven by a Tannoy loudspeaker powered by a power amplifier (B&K 2713) connected to a maximum length sequence system analyzer (MLSSA) with a MLS card installed in a PC. The MLSSA system was used as a source signal generator as well as a signal processing analyzer for all the experimental measurements. The acoustic signals were taken by a BSWA TECH MK224 $\frac{1}{2}$-in. condenser microphone coupled with a BSWA TECH MA201 preamplifier that was connected to the MLS card through a BSWA TECH MC102 amplifier.

B. Characterization and acoustic measurement of a sphere in a free field

To confirm that the theoretical models introduced in Secs. II and III are competent for studying acoustic scattering by an absorbing sphere, a set of experimental measurements was carried out for comparison with the theoretical predictions. A foam sphere with a radius of 0.1 m and density of 18.4 $\mathrm{kg\,m^{-3}}$, which was made of open-cell polyurethane, was used in the experiments of the current study. Beginning with the in-depth experimental validations, preliminary free field measurements were conducted to characterize the acoustical properties of the foam sphere. The sphere was suspended in an anechoic chamber by means of two small hooks bolted at the tips of the sphere and fixed in position using fine threads that were tightly fastened to two poles erected remotely from the sphere as shown in Fig. 4. A specifically designed device, calibrated with angles and radial dimensions, was used to measure the positions of both the source and receiver points.

The acoustic impedance of the foam sphere was determined by comparing the experimental data with the theoretical predictions at a given source/receiver configuration, cf. Eq. (2a). Because the polyurethane foam used for the sphere is relatively dense, we assume that the acoustical impedance can be modeled as a locally reacting type. It is well established that a two-parameter model developed by Attenborough[24] is generally applicable to calculate the acoustic impedance of many locally reacting materials. The specific normalized admittance of the foam sphere is calculated by

$$\beta_s = \frac{1}{0.436(1+i)(\sigma_e/f)^{1/2} + 19.48 i \alpha_e / f}, \tag{29}$$

where f is the frequency, σ_e is the effective flow resistivity, and α_e is the effective rate of change of porosity with depth of the material. Knowing the parametric values of σ_e and α_e, we can calculate β_s and apply Eqs. (2a), (4), and (8) to compute the corresponding sound field at a given geometrical configuration. By adjusting the parametric values of the Attenborough model, we can deduce a set of σ_e and α_e such that the predicted sound levels give a best fit curve with the experimental data.

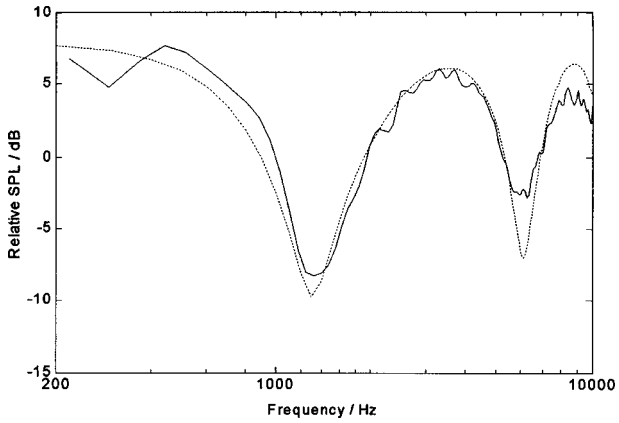


FIG. 7. Characterization of a thick carpet. Source height=0.16 m, receiver height=0.16 m, and range=1 m. The best-fit parameters are found to be $\sigma_e = 10\ \mathrm{kPa\,s\,m^{-2}}$ and $\alpha_e = 80\ \mathrm{m^{-1}}$. (Dotted line: theoretical prediction by the Weyl-van der Pol model and the two-parameter model; solid line: experiment result).

The acoustic characterization of the polyurethane sphere was implemented for different source/receiver configurations. The best-fit parameters were found to be $\sigma_e = 100\ \mathrm{kPa\,s\,m^{-2}}$ and $\alpha_e = 10\ \mathrm{m^{-1}}$ (see Fig. 5 for some typical characterization results). In the plots, we display the spectra of relative sound pressure levels with the source, S, located at $R_o = 0.15$ m, $\theta_o = 90°$, $\varphi_o = 30°$, and the receiver, P, at the following six locations: (a) $R = 0.15$ m, $\theta = 90°$, $\varphi = 30°$; (b) $R = 0.15$ m, $\theta = 90°$, $\varphi = 60°$; (c) $R = 0.15$ m, $\theta = 90°$, $\varphi = 90°$; (d) $R = 0.3$ m, $\theta = 90°$, $\varphi = 120°$; (e) $R = 0.15$ m, $\theta = 90°$, $\varphi = 150°$; and (f) $R = 0.3$ m, $\theta = 90°$, $\varphi = 180°$. It is not surprising to see that the experimental results can generally be predicted by the theoretical calculations. In fact, we have conducted measurements at various source/receiver configurations, and good agreements between theoretical predictions and experimental data can generally be achieved. For the sake of brevity, they have not been shown. As expected, the frequency spectra become rather flat as the subtended angle between the source and receiver increases. This is because the interference effect of the direct wave on the scattered sound at the field points becomes less effective when the subtended angle increases. In such a situation, the receiver is located in the shadow zone. With these enlightening results, we feel confident about extending the theoretical model for the prediction of acoustic diffraction by an absorbing sphere in the presence of a reflecting ground.

C. Acoustic diffraction by a locally reacting sphere above a hard ground

In this set of experiments, the foam sphere was suspended above a hard ground in the same way as described in the previous section. A varnished hardwood board was used to simulate a hard ground. The center of the sphere was set at 0.275 m above the hard ground. In Fig. 6, the spectra of relative sound pressure levels were obtained with the source situated at $R_o = 0.4$ m, $\theta_o = 90°$, $\varphi_o = 0°$ and receivers at (a) $R = 0.15$ m, $\theta = 90°$, $\varphi = 30°$; (b) $R = 0.15$ m, $\theta = 90°$, $\varphi = 60°$; (c) $R = 0.25$ m, $\theta = 90°$, $\varphi = 90°$; (d) $R = 0.25$ m, $\theta = 90°$, $\varphi = 120°$; (e) $R = 0.25$ m, $\theta = 90°$, $\varphi = 150°$; and (f) R

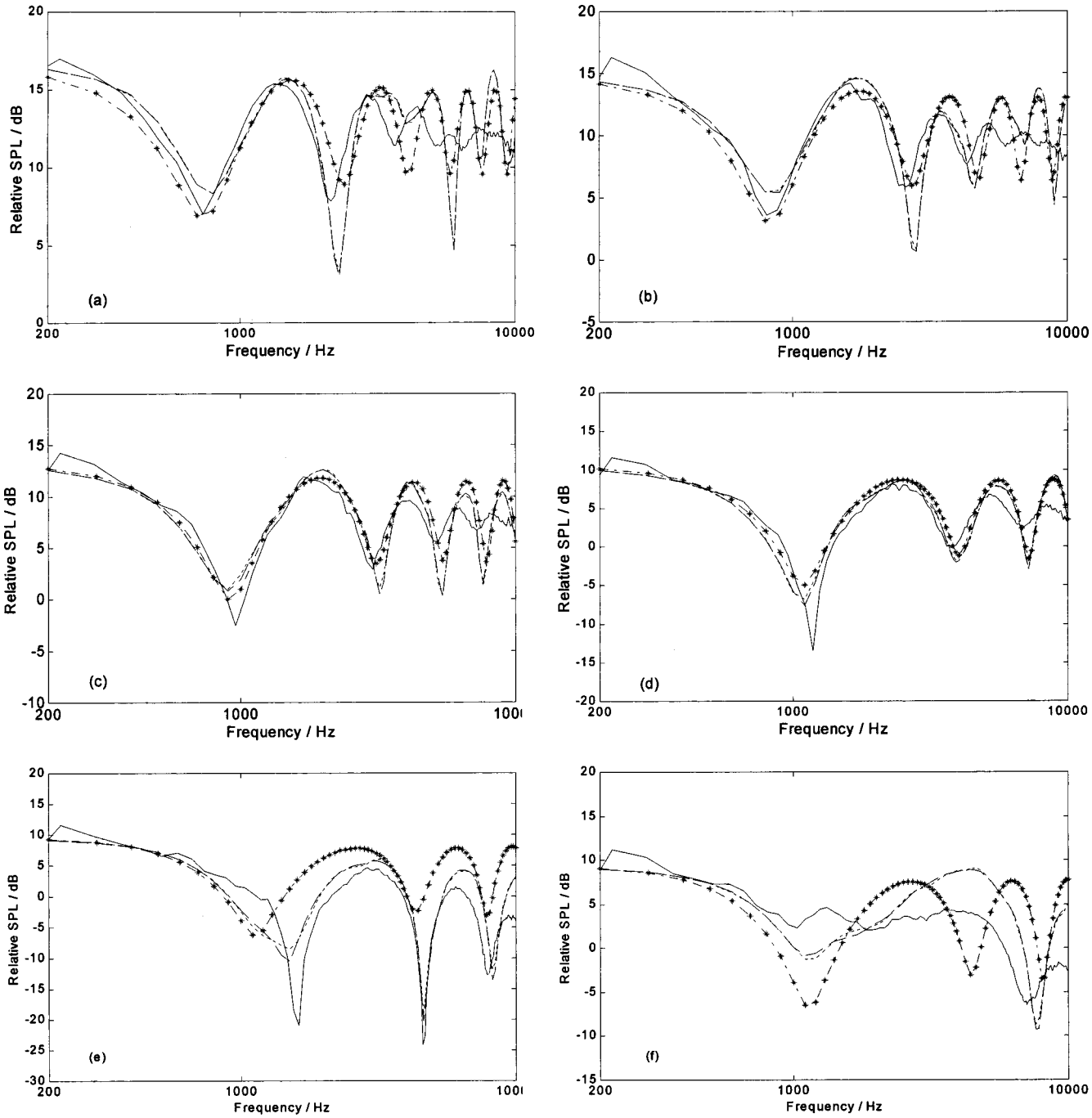


FIG. 8. Experimental results and theoretical predictions for acoustic diffraction by a locally reacting sphere with a radius of 0.1 m above an impedance ground. The center of the sphere is 0.176 m above the impedance ground. Source: $R_o=0.35$ m, $\theta_o=90°$, $\varphi_o=0°$. Receiver: (a) $R=0.15$ m, $\theta=90°$, $\varphi=30°$; (b) $R=0.15$ m, $\theta=90°$, $\varphi=60°$; (c) $R=0.15$ m, $\theta=90°$, $\varphi=90°$; (d) $R=0.30$ m, $\theta=90°$, $\varphi=120°$; (e) $R=0.30$ m, $\theta=90°$, $\varphi=150°$; and (f) $R=0.30$ m, $\theta=90°$, $\varphi=180°$. (Dotted line: theoretical prediction—approximation; dashed line: theoretical prediction—exact solution; solid line: experimental result; dash-dotted line with legend: predicted sound field without presence of sphere.)

$=0.25$ m, $\theta=90°$, $\varphi=180°$. In these figures, the experimental results are plotted in solid lines. The dashed lines represent the theoretical predictions derived from an exact solution method by solving Eq. (17), and the dotted lines represent the theoretical result obtained by an approximate method using Eq. (22). The predicted results from these two methods are very close to within the thickness of the curves shown in the figures. The predicted sound field (dash-dotted line) at each receiver position above a hard ground in the absence of the sphere is also given in the figures. They serve to show how the sphere perturbs the sound field above a hard plane. In the numerical analyses, we have used $N=30$ [the number of terms required for the summation series in Eq. (16)] throughout in order to obtain the numerical results calculated by using an exact solution method. We can see that the experimental results show close agreement with the predictions based on the analytical models discussed in Sec. III. However, the computational time needed to obtain the same prediction result by using the approximation solution method is much faster than by using the exact solution method. We

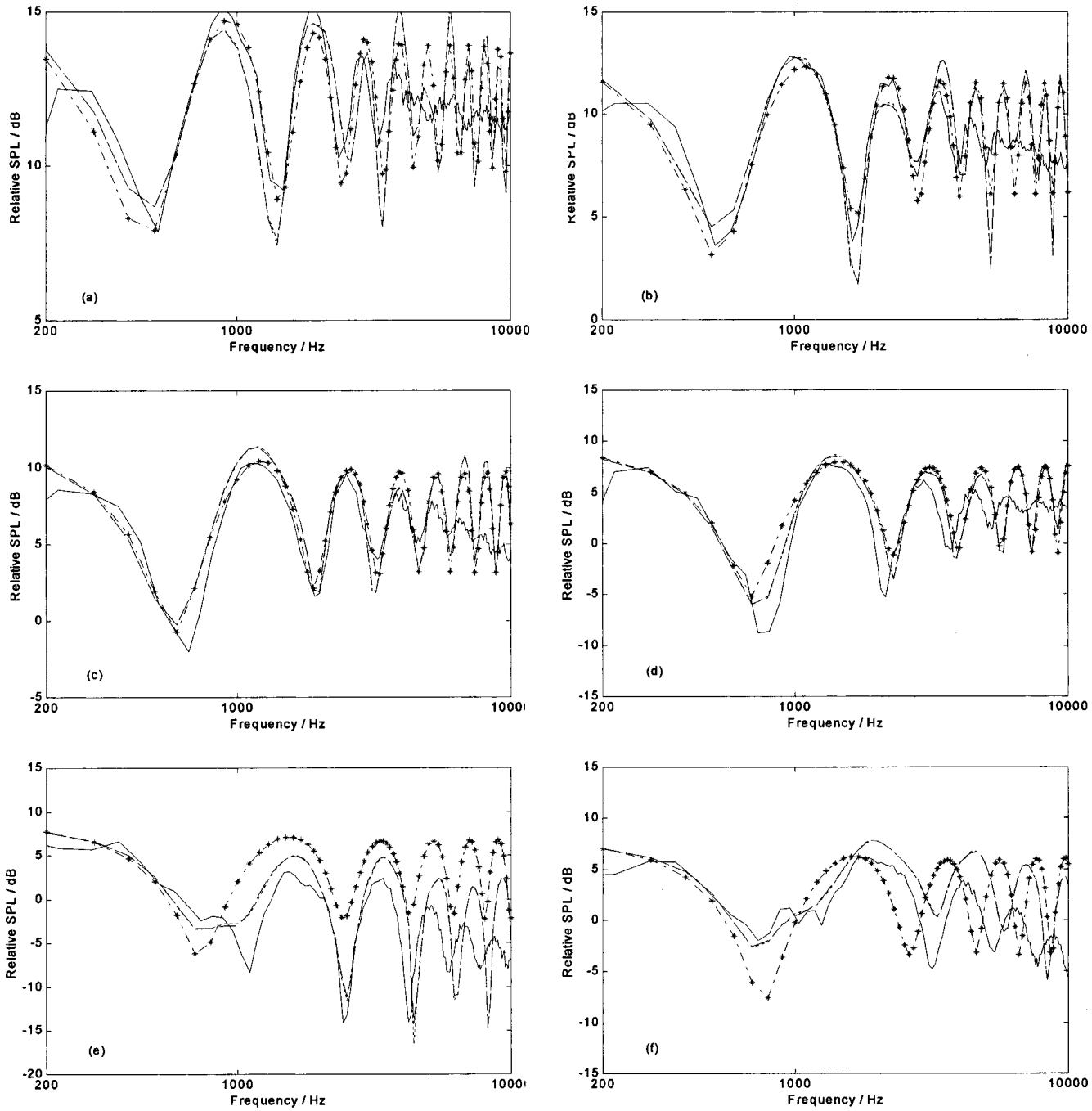


FIG. 9. Experimental results and theoretical predictions for acoustic diffraction by a locally reacting sphere with a radius of 0.1 m above an impedance ground. The center of the sphere is 0.26 m above the impedance ground. Source: $R_o=0.4$ m, $\theta_o=90°$, $\varphi_o=0°$. Receiver: (a) $R=0.2$ m, $\theta=90°$, $\varphi=30°$; (b) $R=0.2$ m, $\theta=90°$, $\varphi=60°$; (c) $R=0.2$ m, $\theta=90°$, $\varphi=90°$; (d) $R=0.3$ m, $\theta=90°$, $\varphi=120°$; (e) $R=0.30$ m, $\theta=90°$, $\varphi=150°$; and (f) $R=0.35$ m, $\theta=90°$, $\varphi=180°$. (Dotted line: theoretical prediction—approximation; dashed line: theoretical prediction—exact solution; solid line: experimental result; dash-dotted line with legend: predicted sound field without presence of sphere.)

developed a MATLAB program for each model and used a desktop computer with CPU of 1.8 GHz to carry out the computations. The program takes 1 s for the approximate model compared with 60 s for the exact solution to obtain a predicted result at a source frequency of 1000 Hz. Understandably, the interference patterns of the spectra are more intricate in this case, as the reflected waves can interact with both the direct and scattered waves to create a more complicated sound field even when the measuring point is in the shadow zone. Again, measurements at other source/receiver configurations and their corresponding theoretical predictions are not shown for brevity, but they all show good agreement with each other.

D. Acoustic diffraction by a locally reacting sphere above an impedance ground

The impedance ground was simulated by laying a thick carpet on a hardwood board. The carpet can be regarded as a locally reacting ground, and a two-parameter model applying

Eq. (29) was used to find the effective admittance of the ground. Prior measurements were conducted to characterize the acoustical properties of the carpet. Figure 7 shows the result of theoretical prediction using the Weyl–van der Pol model applying Eq. (23) and the experimental result for characterizing the impedance (admittance) of the carpet. This measurement result was subsequently fitted with the two-parameter model. The best-fit parameters were found to be $\sigma_e = 10\ \mathrm{kPa\,s\,m^{-2}}$ and $\alpha_e = 80\ \mathrm{m^{-1}}$. These acoustical parameters were used in our subsequent predictions of acoustic diffraction by an absorbing sphere above an impedance ground.

The foam sphere was suspended above the impedance ground in the same way as described in the previous sections. Two sets of experiments were carried out in this case. In the first set of measurements, the center of the sphere was set at 0.176 m above the impedance ground; for the second set of experiments, at 0.26 m above the ground. In the first set of measurements (Fig. 8), the spectra of relative sound pressure levels were obtained with the source at $R_o = 0.35\ \mathrm{m}$, $\theta_o = 90°$, $\varphi_o = 0°$ and receivers at (a) $R = 0.15\ \mathrm{m}$, $\theta = 90°$, $\varphi = 30°$; (b) $R = 0.15\ \mathrm{m}$, $\theta = 90°$, $\varphi = 60°$; (c) $R = 0.15\ \mathrm{m}$, $\theta = 90°$, $\varphi = 90°$; (d) $R = 0.30\ \mathrm{m}$, $\theta = 90°$, $\varphi = 120°$; (e) $R = 0.30\ \mathrm{m}$, $\theta = 90°$, $\varphi = 150°$; and (f) $R = 0.30\ \mathrm{m}$, $\theta = 90°$, $\varphi = 180°$. Figure 9 shows the results of the second set of measurements. The spectra of relative sound pressure levels were taken with the source at $R_o = 0.4\ \mathrm{m}$, $\theta_o = 90°$, $\varphi_o = 0°$ and receivers at (a) $R = 0.2\ \mathrm{m}$, $\theta = 90°$, $\varphi = 30°$; (b) $R = 0.2\ \mathrm{m}$, $\theta = 90°$, $\varphi = 60°$; (c) $R = 0.2\ \mathrm{m}$, $\theta = 90°$, $\varphi = 90°$; (d) $R = 0.3\ \mathrm{m}$, $\theta = 90°$, $\varphi = 120°$; (e) $R = 0.30\ \mathrm{m}$, $\theta = 90°$, $\varphi = 150°$; and (f) $R = 0.35\ \mathrm{m}$, $\theta = 90°$, $\varphi = 180°$. In these figures, the experimental results are plotted in solid lines. The dashed lines represent the theoretical predictions derived from an exact solution method, and the dotted lines represent the theoretical results obtained by the approximation scheme. The computed results from both prediction methods agree to within the thickness of the lines shown in the figures. We also remark that the predicted sound field in the absence of the sphere above an impedance plane is also shown (dash-dotted lines) in Figs. 8 and 9. These predictions are shown for assessing the relative importance of the contribution from the ground reflected waves as compared with those scattered by the sphere.

The predictions made by the theoretical model developed in Sec. III B agree reasonably well with the experimental results. Not surprisingly, the theoretical predictions are in accord with the patterns observed from experimental measurements in spite of the fact that the experimental results show a trend of noticeable differences in magnitude (the experimental results are always lower than predictions) at high frequency region. These discrepancies are mainly due to experimental limitations, especially the production of a point source by a long tube. The point source at high frequency may not be omni-directional and the energy content of the source may not be sufficiently higher than the background noise level. Similar to the case of sound diffraction near a hard ground, the interference patterns of the spectra are more intricate compared to the sound diffraction by an absorbing sphere in a free field as the total sound field is made up of reflected waves, direct waves, and scattered waves to give a more complicated acoustic environment.

V. CONCLUSIONS

The theoretical formulations of acoustic diffraction by an absorbing sphere of different boundary conditions insonified by a point source have been discussed. The acoustical properties of the sphere made of open-cell polyurethane foam have been characterized by indoor measurements and used *a posteriori* to validate the theoretical models. The theoretical analyses have been extended to the study of acoustic diffraction by an absorbing sphere of different boundary conditions above an acoustically hard and an impedance ground. Exact analytical solutions have been derived using the classical method of images and translational addition theorems for the spherical wave functions. The effect of the impedance surface on the ground-reflected sound fields can be incorporated into the exact analytical model through the use of the Weyl–van der Pol formula, which is an approximate analytical solution for the reflection of a spherical wave from a locally reacting plane. The analytical solutions developed for the presence of a hard or an impedance ground can be expressed in four terms: contributions from the source and the sphere, and contributions from the image source and the image sphere. An approximate method has been proposed to obtain the scattering coefficients for a sphere located not too close to the ground, which can substantially reduce the computational time for calculating the total sound fields. Experimental results have been used to validate the theoretical predictions made by the analytical models and good agreement was observed. It has also been shown that the numerical result predicted by the approximate method agrees closely with the prediction made by the exact solution method. Not surprisingly, the numerical simulations and experimental results reveal that the sound fields become more intricate in the presence of a reflecting ground. The results of the current study can serve as a benchmark for solutions obtained by other numerical or asymptotic approaches.

ACKNOWLEDGMENTS

The research described in this paper was financed jointly by the Innovation and Technology Commission of the Hong Kong Special Administrative Region and the Mass Transition Railway Corporation Limited, through the award of an Innovation and Technology Fund under the category of the University-Industry Collaboration Program (Project No. UIM/39). The project was supported in part by the Research Grants Committee of the Hong Kong Special Administrative Region. The authors gratefully acknowledge the technical and administrative support given by The Hong Kong Polytechnic University.

APPENDIX: RECURSION RELATION FOR VERTICAL TRANSLATION OF SPHERICAL WAVE FUNCTIONS

When the translation for the spherical wave functions is along the vertical z axis, the Gaunt coefficients can be evaluated by a recursion relation which is highly desirable, espe-

cially for machine computation, as there is no need to calculate any Wigner $3-j$ symbol. The recursion relation is given by

$$\alpha_{p-3}a_{p-4}-(\alpha_{p-2}+\alpha_{p-1}-4m^2)a_{p-2}+\alpha_p a_p=0, \tag{A1}$$

where

$$a_p \equiv a(m,n,-m,\nu,p), \tag{A2}$$

$$p=n+\nu,n+\nu-2,...,|n-\nu|, \tag{A3}$$

and

$$\alpha_p=\frac{[(n+\nu+1)^2-p^2][p^2-(n-\nu)^2]}{4p^2-1}. \tag{A4}$$

The recursion relation is most conveniently employed in the backward direction, since we can find simple starting values for the coefficients. The first two values at $p=n+\nu$ and $p=n+\nu-2$ are

$$a_{n+\nu}=\frac{(2n-1)!!(2\nu-1)!!}{(2n+2\nu-1)!!}\frac{(n+\nu)!}{(n-m)!(\nu+m)!}, \tag{A5}$$

$$a_{n+\nu-2}=\frac{(2n+2\nu-3)}{(2n-1)(2\nu-1)(n+\nu)} \times[n\nu-m^2(2n+2\nu-1)]a_{n+\nu}, \tag{A6}$$

where

$$(2q-1)!!=(2q-1)(2q-3)\cdots 3\cdot 1;$$

$$(0)!!\equiv(-1)!!\equiv 1. \tag{A7}$$

This recursion relation can give all a_p coefficients and thus all $A_{m\nu}^{mn}$.

[1] G. C. Gaunaurd and H. Huang, "Acoustic scattering by a spherical body near a plane boundary," J. Acoust. Soc. Am. **96**, 2526–2536 (1994).
[2] G. C. Gaunaurd and H. Huang, "Sound scattering by a spherical object near a hard flat bottom," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 690–697 (1996).
[3] G. C. Gaunaurd and H. Huang, "Acoustic scattering by an air-bubble near the sea surface," IEEE J. Ocean. Eng. **20**(4), 285–292 (1995).
[4] H. Huang and G. C. Gaunaurd, "Acoustic point source scattering by a spherical elastic shell submerged beneath a free surface," J. Acoust. Soc. Am. **99**, 2720–2726 (1996).
[5] E. L. Shenderov, "Diffraction of sound by an elastic or impedance sphere located near an impedance or elastic boundary of a halfspace," Acoust. Phys. **48**(5), 607–617 (2002).
[6] G. Dassios, M. Hadjinicolaou, and G. Kamvyssas, "The penetrable coated sphere embedded in a point source excitation field," Wave Motion **32**, 319–338 (2000).
[7] D. S. Jones, *Acoustic and electromagnetic waves* (Oxford U. P., Oxford, 1986), pp. 441–442. (Note that in this textbook, the time dependent factor $e^{i\omega t}$ is used.)
[8] F. M. Wiener, "Sound diffraction by rigid spheres and circular cylinders," J. Acoust. Soc. Am. **19**(3), 444–451 (1947).
[9] L. D. Hampton and C. M. McKinney, "Experimental study of the scattering of acoustic energy from solid metal spheres in water," J. Acoust. Soc. Am. **33**(5), 664–673 (1961).
[10] M. L. Harbold and B. N. Steinberg, "Direct experimental verification of creeping waves," J. Acoust. Soc. Am. **45**(3), 592–603 (1961).
[11] W. K. Lui and K. M. Li, "A theoretical study of the horn amplification of sound radiated from tires above a porous road pavement," J. Acoust. Soc. Am. (Submitted, 2003).
[12] A. D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications* (Acoustical Society of America, New York, 1989), Sec. 9.8.
[13] S. Stein, "Addition theorems for spherical wave functions," Q. Appl. Math. **19**, 15–24 (1961).
[14] O. R. Cruzan, "Translational addition theorems for spherical vector wave functions," Q. Appl. Math. **20**, 33–40 (1962).
[15] J. H. Bruning, "Multiple scattering by spheres," Ph.D. thesis, University of Illinois, 1969.
[16] J. H. Bruning and Y. Lo, "Multiple scattering of electromagnetic waves by spheres I," IEEE Trans. Antennas Propag. **AP-19**, 378–400 (1971).
[17] C. F. Chien and W. W. Soroka, "A note on the calculation of sound propagation along an impedance plane," J. Sound Vib. **69**, 340–343 (1980).
[18] R. K. Pirinchieva, "Model study of sound propagation over ground of finite impedance," J. Acoust. Soc. Am. **90**(5), 1191–2682 (1991).
[19] R. K. Pirinchieva, Erratum: "Model study of sound propagation over ground of finite impedance," J. Acoust. Soc. Am. **94**(3), 1722 (1993).
[20] M. E. Delaney and E. N. Bazley, "Acoustical properties of fibrous materials," Appl. Acoust. **3**, 105–116 (1970).
[21] T. F. W. Embleton, "Tutorial on sound propagation outdoors," J. Acoust. Soc. Am. **100**, 31–48 (1996).
[22] A. Edmunds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1957).
[23] S. M. Hasheminejad, "Modal acoustic force on a spherical radiator in an acoustic halfspace with locally reacting boundary," Acta Acust. **87**, 443–453 (2001).
[24] K. Attenborough, "Ground parameter information for propagation modeling," J. Acoust. Soc. Am. **92**, 418–427 (1992).

Measurement of six acoustical properties of a three-layered medium using resonant frequencies

Carlos L. Yapura[a)]
G Systems, Inc., Plano, Texas 75074

Vikram K. Kinra[b)] and Konstantin Maslov[c)]
Department of Aerospace Engineering, Texas A&M University, College Station, Texas 77845-3141

(Received 15 April 2003; revised 15 September 2003; accepted 3 October 2003)

An ultrasonic technique for a simultaneous determination of traveling times and acoustic impedances of three-layered, platelike specimens using a normally incident longitudinal wave is presented. The acoustical properties are determined by minimizing the difference between measured and calculated resonant frequencies in the well-known least-squares sense. The resonant frequencies of the layered structure are deduced, in transmission, from the maxima of the magnitude of the transmission spectrum, or, in reflection, from the π values of the phase of the reflection spectrum. Measurements are carried out in the frequency range 1–20 MHz for steel/aluminum/steel and silver/copper/nickel specimens with individual layer thickness values ranging from 0.3 mm to 2 mm. The differences between the inverted thicknesses and those measured directly with an optical microscope range from -4 μm to $+13$ μm. Estimated error bounds suggest that the inverted travel times are accurate within ± 1 ns with 95% probability. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1630998]

PACS numbers: 43.20.−f, 43.20.Ye, 43.20.Ks [DEC] Pages: 57–65

I. INTRODUCTION

Multilayered media are being used in a variety of engineering applications such as protective coatings, lap joints, laminated and filament wound composites, and clad-metal structures. Accordingly, it is necessary to measure certain engineering properties (for example, thickness and stiffness of the individual layers) for quality assurance or for detection of structural degradation due to wear and adverse environmental conditions.[1] Some ultrasonic NDE (non-destructive evaluation) examples include the evaluation of an adhesive bond,[2,3] the thickness measurement of a thin coating on a thick substrate,[4,5] and the acoustic microscopy of a layer on a thick substrate.[6,7] A limitation of most of the existing ultrasonic NDE techniques is that to measure the acoustical properties of a single layer the properties for all other layers must be known. However in many applications the properties of several layers are unknown. Therefore, a multiparameter inversion must be carried out on a multi-layered specimen even when the properties of only one layer are needed.

A multi-parameter inversion for a multi-layered specimen using the resonant scattering of longitudinal waves due to excitation of guided waves in the solid layers was carried out by Lenoir *et al.*[8] However, in their work the presence of a liquid layer between the solid layers was critical for the measurement of longitudinal velocities in the solid layers. Moreover, the incident ultrasonic pulses had to be short compared to the travel time in each layer to make at least the first three reflections from the plate surfaces clearly separable in the time domain. Kinra *et al.*[9] used the transmitted complex spectrum of normally incident longitudinal waves to successfully invert within $\pm 5\%$ for three thicknesses and one wavespeed of three-layered specimens composed of metal-water-metal layers even when the reflection from the boundaries of the layers were nonseparable.

The objective of this paper is to invert successfully the maximum possible number of acoustical properties of a multilayered media, for example, thicknesses and wave speeds of three-layered specimens using through-thickness resonant frequencies of waves measured in transmission as well as in reflection. These resonant frequencies do not have a strong dependence on attenuation, beam spreading, and angle of incidence compared to a signal magnitude (see exact solution in Lobkis *et al.*[10]). The other essential goal is to find an estimate for the error bounds for the inverted parameters for a given level of confidence.

II. EXPERIMENTAL SETUP

A schematic of the experimental setup used for transmission and reflection measurements is shown in Fig. 1. It is described in detail by Kinra and Dayal[11] and Kinra *et al.*[9] In transmission, a three layer sample was placed in the near field zone between an aligned pair of 12.7 mm diameter ultrasonic transducers. A Panametrics 5052U ultrasonic analyzer was used as pulser/receiver. The transmitted signal through the layered media was digitized into 16 384 points by a Tektronix DSA 601 digital signal analyzer with 8-bit precision at a sampling rate of 1 GHz.

The transfer function of the specimen was calculated from the ratio of the FFTs of the digitized transmitted and reference signals. The reference signal was obtained by measuring the signal across the immersion liquid with the speci-

[a)]Project Engineer.
[b)]General Dynamics Chair Professor.
[c)]Author to whom correspondence should be addressed. Electronic mail: maslov@tamu.edu

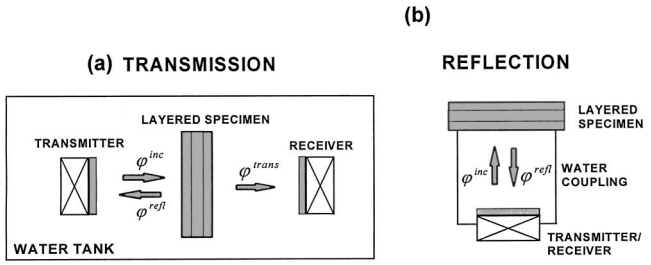


FIG. 1. Schematic of the experimental setup used for (a) transmission and (b) reflection measurements.

men removed. In reflection, the specimen was placed at the open end of a water-filled chamber opposite to a 12.7 mm diameter piezoelectric transducer. A reference signal was obtained by placing the front surface of a ''thick'' block of aluminum at the same location as the front face of the layered specimen. The systematic error in the measurement of the phase of the transfer function in reflection due to an uncertainty in positioning the reference block is estimated from the resulting travel-time shift, τ_S (see Secs. III B and IV B). This value of τ_S is used to correct the phase of the reference signal by multiplying its FFT by the factor $\exp(i\omega\tau_S)$.

Three metal-clad specimens were evaluated in this work: 304SS-Al-304SS, 436SS-Al-304SS, and Ag-Cu-Ni. The stainless steel/3004 aluminum/stainless steel samples were manufactured by Clad Metals, Inc.[12] The silver/copper/nickel sample was manufactured by Polymetallurgical Corporation.[13] The material and geometrical properties of each layer are listed in Table I. The densities of 304SS and 436SS were obtained from single layer samples using Archimedes' Principle (accuracy $\pm 0.2\%$), and the wavespeeds using a procedure reported by Kinra and Iyer[14] (accuracy $\pm 0.5\%$). The density and wavespeed of aluminum are taken from Kinra and Iyer.[14] The densities and wavespeeds of silver, copper, and nickel are taken from Ensminger.[15] At a temperature of 22 °C, the density and wavespeed of water are taken to be 0.997 g/cc 1.49 mm/μs, respectively.[14] The thicknesses of the individual layers are also listed in Table I. Each thickness was measured using an optical microscope equipped with a micrometer-driven, positioning stage. The specimens were polished on one side and ten readings were taken along the sample. The uncertainty in the measurement of each thickness ranges from ± 1 μm to ± 4 μm.

TABLE I. Expected values for the acoustical properties.

Property	304SS/Al/304SS	436SS/Al/304SS	Ag/Cu/Ni
ρ_1 (g/cc)	7.93	7.73	10.5
ρ_2 (g/cc)	2.71	2.71	8.90
ρ_3 (g/cc)	7.93	7.93	8.80
c_1 (mm/μs)	5.76	6.10	3.70
c_2 (mm/μs)	6.40	6.40	4.66
c_3 (mm/μs)	5.76	5.76	5.63
h_1 (mm$\pm\mu$m)	0.375±2	0.438±2	0.344±3
h_2 (mm$\pm\mu$m)	1.530±3	1.833±3	1.132±4
h_3 (mm$\pm\mu$m)	0.375±3	0.388±1	0.285±2
h (mm$\pm\mu$m)	2.279±2	2.662±2	1.766±4

III. FORWARD SOLUTION

A. Time domain: transmission

Consider the propagation of a short-duration longitudinal wave in a fully immersed layered medium as seen in Fig. 1. In the time domain, the signal transmitted through the layered sample consists first of a direct (through-pass) transmitted pulse and multiple echoes from the boundaries of the layers. In an ideally elastic nondispersive medium even after all these multiple passes, the shape of the incident pulse (wavelet) is preserved. However, its amplitude changes in accordance with the transmission and reflection coefficients at the boundaries it crosses or reflects from. Furthermore, these wavelets have different delay times corresponding to the travel paths they take. Accordingly, the first through-transmitted wavelet, $g(t)$, can be calculated from the measured reference wavelet $f(t)$ using the following set of equations:

$$g(t)=Tf(t+\tau), \tag{1}$$

$$\tau=(h_1+h_2+h_3)/c_0-(\tau_1+\tau_2+\tau_3), \tag{2}$$

$$T=T_{01}T_{12}T_{23}T_{34}, \tag{3}$$

$$T_{j-1,j}=2/(1+r_j), \quad h_j=\tau_j c_j=\tau_j z_j/\rho_j,$$

$$r_j=z_j/z_{j-1}, \quad z_j=\rho_j c_j, \tag{4}$$

where τ is the travel time through a three-layer specimen subtracted from the travel time through a water layer of the same total thickness, T is the one-pass transmission coefficient across the immersed specimen and z_j, τ_j, ρ_j, h_j, and c_j, are the acoustic impedance, longitudinal wave travel time, density, thickness and longitudinal wave speed of the jth layer, respectively. The acoustic impedance ratio r_j is between the jth layer or half-space and the layer (or half-space) indexed before it. The indices, 1, 2, and 3 refer to the individual layers, and indices ''0'' and ''4'' refer to the two ''half-spaces'' of water.

B. Time domain: reflection

In reflection the virgin wave $g(t)$ can be expressed via the reference wavelet $f(t)$, measured as a reflected signal from a ''thick'' reference specimen, as follows:

$$g(t)=\frac{R_{01}}{R_{01}^r}f(t+\tau_s), \tag{5}$$

where τ_s is a time shift due to the experimental error from positioning the front surfaces of the reference and the layered media, R_{01}^r is the reflection coefficient of the reference specimen of known properties and R_{01} is the reflection coefficient from the boundary between water and first sample layer,

$$R_{01}=\frac{r_1-1}{r_1+1}. \tag{6}$$

C. Frequency domain: transmission

The transfer function, H_T^*, is defined as the deconvolution of a received ultrasonic signal with respect to a reference signal for a three-layered medium. For a three-layered medium, for which experiments are carried out in this work, H_T^* is given by[16]

$$H_T^* = 2^4 e^{i\omega\tau}[(1+r_1)(1+r_2)(1+r_3)(1+r)+(1-r_1)(1+r_2)(1+r_3)(1-r)e^{-2i\omega(\tau_1+\tau_2+\tau_3)}+(1-r_1)(1-r_2)(1+r_3)\times(1+r)e^{-2i\omega\tau_1}+(1+r_1)(1-r_2)(1+r_3)(1-r)e^{-2i\omega(\tau_2+\tau_3)}+(1+r_1)(1-r_2)(1-r_3)(1+r)e^{-2i\omega\tau_2}+(1-r_1)\times(1+r_2)(1-r_3)(1+r)e^{-2i\omega(\tau_1+\tau_2)}+(1+r_1)(1+r_2)(1-r_3)(1-r)e^{-2i\omega\tau_3}+(1-r_1)(1-r_2)(1-r_3)\times(1-r)e^{-2i\omega(\tau_1+\tau_3)}]^{-1}, \quad (7)$$

where ω is the circular frequency and $r=r_4=1/(r_1r_2r_3)$. The fluid-loaded resonant frequencies are defined as the frequencies where the magnitude of the transfer function is a local maximum, i.e., the first derivative of the magnitude with respect to the frequency is zero and the second derivative with respect to the frequency is negative. The fluid-loaded resonances are therefore obtained from the roots of the resonant frequency equation

$$F=\partial H_T/\partial\omega=0, \quad (8)$$

where H_T is the magnitude of the complex-valued transfer function. From a careful observation of Eq. (10), note that H_T is a function of τ_1, τ_2, τ_3 and r_1, r_2, r_3 only. If the densities ρ_1, ρ_2, and ρ_3 are assumed to be known, the six parameters can be written as thicknesses and wave speeds, h_1, h_2, h_3, c_1, c_2, and c_3. Extending the observation made of H_T to N-layered media, the total number of parameters that can be inverted using fluid-loaded resonant frequencies is $2N$, that includes N travel times and N impedance ratios.

D. Frequency domain: reflection

More often than not, layered media to be evaluated cannot be accessed from both sides. If the specimen is accessed from one side and the other side is free of tractions, then the measured resonant frequencies are those at which the input impedance of the specimen is equal to zero, which are the same as the resonant frequencies of the specimen *in-vacuo*. For three-layer medium the the *in-vacuo* resonant frequencies are given by[17]

$$F=\sin\omega\tau_1(\cos\omega\tau_2\cos\omega\tau_3-r_3\sin\omega\tau_2\sin\omega\tau_3)+r_2\cos\omega\tau_1(\sin\omega\tau_2\cos\omega\tau_3+r_3\cos\omega\tau_2\sin\omega\tau_3)=0. \quad (9)$$

Note that the resonant frequencies are completely described by the three travel times τ_1, τ_2, τ_3 and only two acoustic impedance ratios r_2 and r_3. Because of this, r_1 cannot be deduced from measurements of the resonant frequencies. Therefore, the number of acoustical parameters that can be obtained using resonant frequencies for an N-layer media with the back surface free is $2N-1$ that includes N travel times and $N-1$ impedance ratios.

IV. MULTIPARAMETER INVERSION

A. Obtaining constraints for the acoustical properties from inversion of acoustical information in the time domain: transmission

Using (1) the maximum likelihood estimate of the properties τ and T are obtained by minimizing the error function

$$E=\int_{t_0}^{t_1}[g^e(t)-Tf(t+\tau)]^2\,dt \Big/ \int_{t_0}^{t_1}[g^e(t)]^2\,dt, \quad (10)$$

that is the mean square difference between the experimentally measured transmitted signal, $g^e(t)$, and that reconstructed from the reference wavelet, $Tf(t+\tau)$, normalized with respect to the energy of the transmitted signal. Integration is performed over the leading edge of the transmitted signal (so-called virgin wave) from the beginning of the one-pass transmitted signal t_0, time when the signal amplitude confidently exceeds the threshold value defined by the noise level, to the beginning of the arrival of the next wavelet, t_1. The duration of the virgin wave is equal to the fastest single layer round-trip travel time of all layers. The downhill simplex method[18] was used to minimize the error function. The optimal value of t_1, is found by minimizing the value of $\hat{E}$, that is the residual value of E at convergence, as a function of t_1. The value of $\hat{E}$ as a function of t_1 is shown in Fig. 2 about the expected time for the second arrival for a 436SS/Al/304SS specimen. Note that $\hat{E}$ has a minimum at $t_1=1208$ ns; therefore, the properties τ and T obtained at this value are the chosen best estimates. These inverted properties are used to reconstruct the virgin wave as shown in Fig. 3. The inverted properties from the virgin wave are listed in Table II for the three specimens considered. The values in parentheses indicate the differences between the inverted values and the expected values calculated using Table I.

The time domain inversion above for τ and/or T effectively yields two/one constraint(s) on the values that the unknown properties may take, effectively reducing by two/one the number of unknown properties of a layered medium. In principle, the first wavelet can be subtracted from the transmitted signal and the foregoing procedure can be repeated until the inversion of all acoustical properties in the time domain is completed.[19] However, the inversion errors in τ and T propagate to the inversion of the next delay time and magnitude making the latter less accurate. Moreover, the

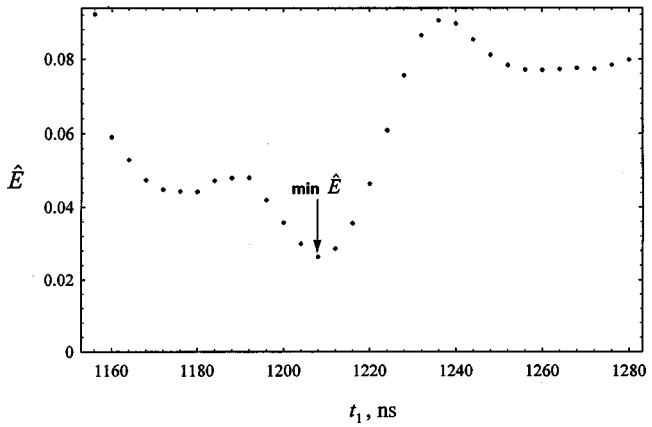


FIG. 2. The residuals values $\sqrt{\hat{E}}$ for the virgin wave as a function of t_1 for 436SS/Al/304SS.

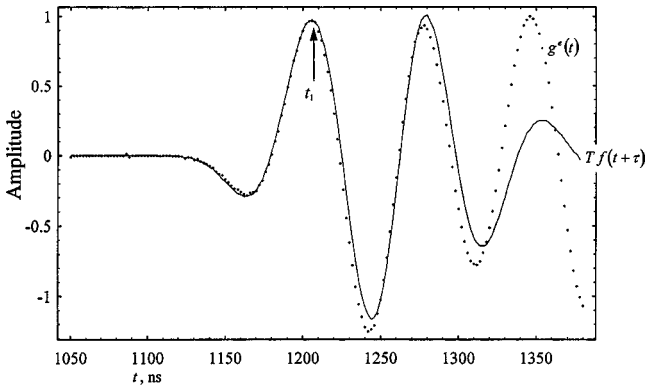


FIG. 3. Reconstruction of the leading edge of the transmitted signal using the incident wave.

chance that an extended part of the ''virgin'' wavelet does not overlap with a subsequent wavelet quickly diminishes with an increasing number of wavelets arriving at about the same delay times due to multiple reflections from the boundaries of the layers. For this reason, only the inversion results for the first wavelet are used in Sec. IV C below.

B. Obtaining a constraint for the acoustical properties from inversion of acoustical information in the time domain: reflection

Using (5) the maximum likelihood estimate of the properties τ_s and R_{01} are obtained by minimizing the error function

$$E=\int_{t_0}^{t_1}\left[g^e(t)-\frac{R_{01}}{R_{01}^r}f(t+\tau_s)\right]^2 dt \Big/ \int_{t_0}^{t_1}[g^e(t)]^2\,dt, \tag{11}$$

where $g^e(t)$ is the measured reflected signal. The inverted value R_{01} can be used to directly find the impedance ratio r_1:

$$r_1=\frac{R_{01}-1}{R_{01}+1}. \tag{12}$$

However, when R_{01} approaches unity (for clad metal samples used in our experiment $R_{01}\sim 0.95$), the relative error in the measurement of r_1 becomes much greater than the error in the measurement of R_{01}. Correspondingly, we cannot use R_{01} to constrain (invert) the value of the acoustical property r_1 for the clad metal specimens.

C. Inversion for the acoustical properties in the frequency domain: transmission

Using Eq. (11) to calculate $f_k(p_\alpha)$, the maximum likelihood estimate of the properties τ_1, τ_2, τ_3, r_1, r_2, and r_3 can be found by minimizing the error function

$$E=\sum_{k=1}^{n}(\Delta f_k)^2, \tag{13}$$

where $\Delta f_k=f_k^e-f_k(p_\alpha)$ is the difference of the measured resonant frequencies f_k^e and the theoretically calculated resonant frequencies $f_k(p_\alpha)$, and n is the number of measured resonant frequencies. The model depends on η properties (six for a three-layered medium), p_α, $\alpha=1,2,...,\eta$. Note that for tracking purposes italic characters are used as indices for the measured and calculated frequencies and greek characters are used as indices for the properties. In the following equations we will use index notation, in which summation over a repeated index is assumed unless otherwise indicated by using capital letters for indexing.

Constraints on the properties from time-domain inverted values for τ and/or T reduces the number of independent properties to be inverted using Eq. (13). Suppose that of the original η properties, only η_1 are independent, p_α with $\alpha=1,2,\ldots,\eta_1$. The remaining properties $p_{\alpha'}$ with $\alpha'=\eta_1+1,\eta_1+2,\ldots,\eta$ are written in terms of the independent properties $p_{\alpha'}=p_{\alpha'}(p_\alpha)$. So the independent properties for a three-layer specimen are now inverted by using the downhill simplex method of minimization by applying the following steps.

(1) Estimate initial values for τ_1, τ_2, τ_3, r_2, and r_3.
(2) Express r_1 as a function of τ_1, τ_2, τ_3, r_2, r_3 and the time-inverted property τ using Eqs. (2) and (4). Alternatively express r_1 as a function of r_2, r_3 and the time-inverted property T using Eqs. (3) and (4). The reason why r_1 is constrained will become apparent by considering the sensitivities of the resonant frequencies to the acoustical properties. Moreover to write r_1 in terms of the remaining acoustical properties and τ requires an *a priori* knowledge of three material parameters, for example, the densities ρ_1, ρ_2, ρ_3.
(3) Using the estimates of the acoustical properties from

TABLE II. Inverted parameters from the virgin wave in transmission.

Property	304SS/Al/304SS	436SS/Al/304SS	Ag/Cu/Ni
τ (μs)	1.153 (-8 ns)	1.370 ($+11$ ns)	0.802 ($+7$ ns)
T	0.1052 ($+8.04\%$)	0.0976 ($+2.46\%$)	0.1233 (-2.24%)

steps (1) and (2), minimize E in (13) using Simplex method until a convergence criteria is satisfied.

D. Inversion for the acoustical properties in the frequency domain: reflection

Using Eq. (12) to calculate $f_k(p_\alpha)$, the maximum likelihood estimate of the properties τ_1, τ_2, τ_3, r_2, and r_3 can be found by minimizing the error function given by Eq. (13). E from Eq. (13) does not depend on r_1 because the model for f_k, see Eq. (10), does not depend on r_1. Therefore the five properties τ_1, τ_2, τ_3, r_2, and r_3 can be estimated directly using the measured frequencies alone and step (2) in Sec. IV C is omitted.

Amplitude information alone cannot be easily used in the frequency domain for the measurements for f_k^e for a specimen with its back surface free of traction. For the case of ideally elastic layers, the magnitude of the reflection coefficient is equal to one at any frequency. For this reason phase information is used. To obtain the correct phase information it is necessary to match precisely, within a micron, the distance from the transducer to the sample surface with that to the reference surface. Since this is virtually impossible we circumvent this problem in the following manner. The resulting travel-time difference between the arrivals of the reference signal and the reflected signal from the specimen is inverted in the time domain as τ_s by minimizing Eq. (11). Using the shifting property for Fourier transforms, the correct phase of the reflected signal for the specimen is obtained by multiplying its FFT by the factor $\exp(-i\omega\tau_s)$.

E. Sensitivities of the resonant frequencies to the acoustical parameters

The accuracy with which the desired acoustical properties can be inverted from frequency measurements depends on the sensitivity of the resonant frequencies to the acoustical parameters of the layered media. A detailed sensitivity analysis of the transfer function in transmission for a single layer was given by Kinra and Iyer,[14] and for three layers by Hanneman and Kinra[16] and Kinra *et al.*[9] A similar analysis is carried out here for the resonant frequencies. The sensitivities of the resonant frequencies to the travel times and acoustic impedance ratios are evaluated at the discrete values for the resonant frequencies f_k. The sensitivity of f_k to a specific parameter p_A is defined as

$$S_{kA} = p_A \frac{\partial f_k}{\partial p_A}. \tag{14}$$

As an example, the sensitivities of the fluid-loaded resonant frequencies for the 436SS/Al/304SS specimen are plotted in Figs. 4(a) and 4(b) using the expected values of the properties. The sensitivities to the travel times [Fig. 4(a)] are about two orders of magnitude higher than the sensitivities to the impedance ratios [Fig. 4(b)]. For clarity, the values of the sensitivities are connected by thin lines. To interpret the values of the sensitivities, consider, for example, the fourth resonant frequency f_4 (at about 5 MHz) where the sensitivity to τ_2 is about -4 MHz and that to r_3 is about $+0.1$ MHz. Suppose that there is a 1% increase in τ_2, then f_4 would

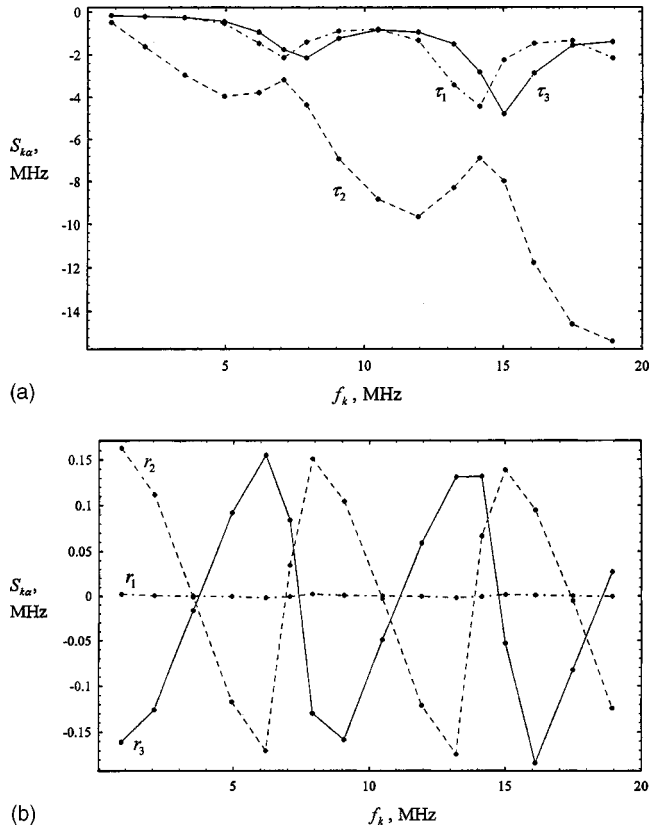


FIG. 4. (a) Sensitivities of the fluid-loaded resonant frequencies to the travel times. (b) Sensitivities of the fluid-loaded resonant frequencies to the acoustic impedance ratios.

decrease by approximately -40 KHz. Similarly, for a 1% increase in r_3, f_4 would increase by only about $+1$ KHz. Conversely, to invert r_3 with a 1% precision using f_4, this resonant frequency would need to be measured accurately within 1 kHz. In the case of water immersion the inversion of r_1 is not possible because the sensitivity of the resonant frequencies to r_1 is very small in the frequency range considered. However, r_1 can be easily obtained using the remaining other five properties and either time-domain inverted property τ or T. The sensitivities of the resonant frequencies *in-vacuo* are very similar to those plotted in Figs. 4(a) and 4(b) and will not be reproduced for brevity.

Additional constraints obtained from inverted values for τ and/or T from time domain inversion reduces the number of independent properties that have to be inverted and consequently the number of sensitivity vectors, S_{kA}. Suppose that of the original η properties, only η_1 are independent, p_α with $\alpha = 1,2,\ldots,\eta_1$. The remaining properties $p_{\alpha'}$ with $\alpha' = \eta_1+1, \eta_1+2,\ldots,\eta$ are written in terms of the independent properties $p_{\alpha'} = p_{\alpha'}(p_\alpha)$. So the sensitivities are now computed using

$$\frac{\partial f_k}{\partial p_\alpha} = \left.\frac{\partial f_k}{\partial p_\alpha}\right|_{\alpha' = \text{const}} + \frac{\partial f_k}{\partial p_{\alpha'}} \frac{\partial p_{\alpha'}}{\partial p_\alpha}, \tag{15}$$

where $\partial p_\alpha / \partial p_{\alpha'}$ is evaluated using the chain rule if the relationships between the properties are implicit.

Once the best estimate p_α^e is obtained from a set of measurements f_k^e, one can obtain a linear approximation for how

this estimate would change given a small change in the measured frequencies. First, consider the following definitions. When a measurement f_k^e is made for a resonant frequency, this measurement lies somewhere about the "true" value $\overline{f_k^e}$ which should be equal to the theoretical value $f_k(\overline{p}_\alpha^e)$. Let the difference between a measurement and its mean (true) value be defined by

$$\Delta f_k^e = f_k^e - \overline{f_k^e}. \tag{16}$$

The inverted properties p_α^e also lie somewhere about the "true" values, $\overline{p}_\alpha^e$. Let this difference be similarly defined by

$$x_\alpha = (p_\alpha^e - \overline{p}_\alpha^e)/\overline{p}_\alpha^e. \tag{17}$$

The reason why the frequencies were not normalized is because it is assumed that the spreads of the measurements are of the same order of magnitude at any frequency. If the spreads were normalized, they would appear large for low resonant frequencies and small for high resonant frequencies.

Consider a minimum of E where the gradient of E with respect to the parameters is zero. Using the definition of E, Eq. (13), this can be written as

$$\partial E/\partial p_A = -\frac{2}{p_A} S_{Ak} \Delta f_k = 0, \tag{18}$$

where S_{Ak} is the sensitivity vector for the parameter p_A, where $A = 1,2...,\eta$. Note that there are n sensitivities for each parameter and, following the convention in this work, summation on k is implied and there is no sum over the repeated capital indices A.

Using Eqs. (16) and (17) and expanding $f_k(p_\alpha)$ linearly in Δf_k about the true values $\overline{p}_\alpha^e$ in Eq. (18) gives

$$S_{\beta k} \Delta f_k^e = S_{\beta k} S_{k\alpha} x_\alpha, \tag{19}$$

that is inverted as

$$x_\alpha = M_{\alpha k} \Delta f_k^e, \tag{20}$$

where

$$M_{\alpha k} = C_{\beta\alpha}^{-1} S_{\beta k}, \quad C_{\beta\alpha} = S_{\beta k} S_{k\alpha}. \tag{21}$$

For a successful inversion it is necessary that the column vectors of $S_{k\alpha}$ be linearly independent. If any column of the sensitivity vector matrix is linearly dependent or if a sensitivity vector is equal to the zero vector, then the determinant of $C_{\alpha\beta}$ is zero and therefore x_α becomes infinitely large. When a sensitivity vector is small compared to the other sensitivity vectors, then the inversion errors for that parameter are large as observed for r_1 during an inversion of all six properties τ_1, τ_2, τ_3, r_1, r_2, and r_3 using the resonant frequencies in transmission. That is why a constraint on r_1 was used in Sec. IV C.

F. Error estimation

The normalized deviations, x_α, were estimated in Eq. (20) for given deviations in the measurements, Δf_k^e. In practice, Δf_k^e are unknown and are errors due to a finite ultrasonic beam size, imperfections on the geometrical shape of the specimen, excitation of Lamb wave modes, and attenuation. A similar problem of parameter inversion of stratified media in time domain was widely discussed elsewhere[20] and it was concluded that, in spite of the complexity of the problem, additive Gaussian noise is a fairly realistic assumption for these errors. Let each measured resonant frequency have a frequency independent additive error with variance σ_f^2. The variance for a measurement set can be expressed, in matrix form, as the covariance matrix

$$\mathrm{cov}(f_k^e) = \sigma_f^2 I_{kl}, \tag{22}$$

where I_{kl} is the identity matrix of dimension equal to the number of measured frequencies, n. Using the linear relationship given by Eq. (20), the covariance matrix for $p_\alpha^e/\overline{p}_\alpha^e$ is expressed as[21]

$$\mathrm{cov}\left(\frac{p_\alpha^e}{\overline{p}_\alpha^e}\right) = M_{\alpha k}\, \mathrm{cov}(f_k^e) M_{\beta l} = \sigma_f^2 C_{\alpha\beta}^{-1}. \tag{23}$$

From Eq. (23), the covariance matrix for the parameters is related proportionally to the variance of the frequency measurements and to the inverse of the "curvature" matrix, $C_{\alpha\beta}$. That is, the sharper the minimum of the error function, the more *precise* the estimate for the unknown properties.

To estimate the errors, expand $f_k(p_\alpha)$ in Eq. (13) linearly near the true values of the properties, $\overline{p}_\alpha^e$. Taking into account that $f_k^e = \overline{f_k^e} + f_k^G$, where f_k^G is the zero-mean Gaussian noise, Eq. (13) yields

$$E = \sum_{k=1}^{n} (f_k^G - S_{k\alpha} x_\alpha)^2. \tag{24}$$

From Eq. (24) one can see that the problem of estimating the error on the properties, x_α, is reduced to a linear regression[21] problem. This can be clearly visualized as follows. Each quantity $y = f_K^G$ is measured at coordinates $s_\alpha = S_{K\alpha}$ in parameter space. The set of constants x_α defines a plane $y = x_\alpha s_\alpha$ passing as close as possible in a least square sense to the measured points along the y direction. Here, a "large" number of measured resonant frequencies takes the place of a "large" number of repeated measurements. Correspondingly, an unbiased estimate for σ_f^2 can be obtained from the residual value of E at convergence,[21]

$$\sigma_f^2 = \hat{E}/(n - \eta_1). \tag{25}$$

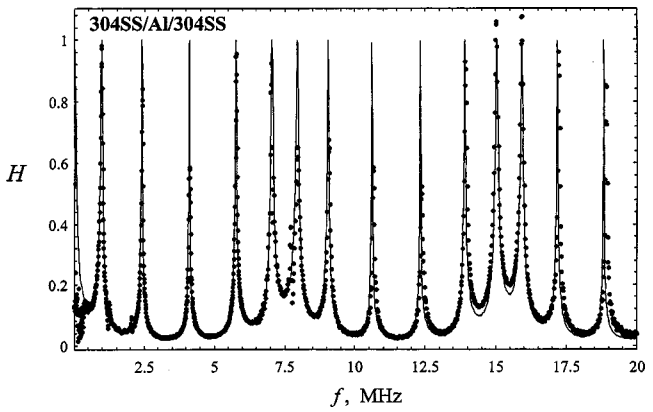


FIG. 5. Magnitude of the transfer function in transmission for 304SS/Al/304SS.

TABLE III. Inversion results for fluid-loaded resonant frequencies using τ.

Property	304SS/Al/304SS	436SS/Al/304SS	Ag/Cu/Ni
$\sqrt{\hat{E}/n}$ (KHz)	8.58	5.21	5.68
τ_1 (μs)	0.0656 (+0.5 ns)	0.0709 (−0.9 ns)	0.0934 (+0.4 ns)
τ_2 (μs)	0.2384 (−0.7 ns)	0.2864 (−0.01 ns)	0.2414 (−1.5 ns)
τ_3 (μs)	0.0656 (+0.5 ns)	0.0676 (+0.2 ns)	0.0514 (+0.8 ns)
r_1	30.16 (−1.88%)	33.10 (+4.35%)	25.79 (−1.36%)
r_2	0.3873 (+2.03%)	0.353 (−4.03%)	1.096 (+2.68%)
r_3	2.582 (−1.99%)	2.618 (−0.62%)	1.183 (−0.97%)
h_1 (mm)	0.371 (−4 μm)	0.451 (+13 μm)	0.341 (−3 μm)
h_2 (mm)	1.527 (−3 μm)	1.836 (+3 μm)	1.139 (+7 μm)
h_3 (mm)	0.371 (−4 μm)	0.388 (−1 μm)	0.290 (+5 μm)
h (mm)	2.269 (−10 μm)	2.674 (+12 μm)	1.770 (+4 μm)

For a single random variable that follows a normal distribution, the probability computed for a given statistical distance gives the percentage that a measurement for that variable will lie within the bounds of the statistical distance. Conversely given a value for the probability or a confidence level, the bounds defined by the resulting statistical distance form a range on a single axis. For two random variables the bounds form an ellipse, for three random variables an ellipsoid, and for four of more random variables a hyperellipsoid. The statistical distance for x_α, $x_\alpha[\mathrm{cov}(x_\alpha)]^{-1}x_\beta$, follows proportionally an F-ratio distribution $F_{\eta_1,n-\eta_1}(1-\gamma)$ with $\eta_1, n-\eta_1$ degrees of freedom.[21] Using this fact, the bounds for a given confidence level γ give the following ellipsoid:

$$x_\alpha C_{\alpha\beta} x_\beta = \frac{\eta_1}{n-\eta_1}\hat{E}F_{\eta_1,n-\eta_1}(1-\gamma), \tag{26}$$

where $0<\gamma<1$. The projection of these ellipsoids into a lesser ($\nu=2$ or $\nu=3$) dimensional space, x_a, where it can be easily visualized is obtained by finding points along the η_1-dimensional ellipsoid where its surface gradient vector is parallel to the hyperplane spanned by all x_a axes. This requires the partial derivatives of (26) with respect to the x_ξ, $\xi=\nu+1,..,\eta_1$, coordinates to be equal to zero. This gives a linear system of equations that is solved for x_ξ in terms of x_a. Then, these solutions are back substituted into (26) giving a quadratic for the ellipse or ellipsoid on the x_a plane. These projections are generalized as follows:

$$x_a C_{ab} x_b = \frac{\eta_1}{n-\eta_1}\hat{E}F_{\eta_1,n-\eta_1}(1-\gamma),$$
$$C^{(\nu)}_{ab} = C_{ab} - C_{a\xi}C^{-1}_{\xi\zeta}C_{\zeta b}, \tag{27}$$

where the projection is made from η_1 dimensions to ν dimensions and $a,b=1,2,\ldots,\nu$ and $\xi,\zeta=\nu+1,\nu+2,\ldots,\eta_1$. For example, the projection into the error axis of the first parameter gives the bounds

$$x_1 = \pm\left[\frac{\eta_1}{n-\eta_1}\hat{E}F_{\eta_1,n-\eta_1}(1-\gamma)/(C_{11}-C_{1\xi}C^{-1}_{\xi\zeta}C_{\zeta 1})\right]^{1/2}. \tag{28}$$

V. INVERSION RESULTS

A. Inversion using transmission data

The first step in the inversion procedure is to select a frequency range containing a suitable number of resonant frequencies. Although six resonances is the minimum number that is required for a least-squares, six-parameter inversion, the accuracy improves considerably by using more resonances [see Eq. (28)]. Moreover, as it can be seen from the sensitivity data in Figs. 4(a) and 4(b) the sensitivity to the travel times increases with frequency. An example of a magnitude spectra H_T used to find resonant frequencies is shown in Fig. 5, corresponding to the 304SS/Al/304SS specimen where the dots represent experimental data and the solid curve is a theoretical curve using Eq. (7).

The inversion procedure for six parameters followed the three steps listed in Sec. IV C. An initial guess for τ_1, τ_2, τ_3, r_2, and r_3 that was $\pm 10\%$ different from the expected values was used. Then, the set was completed by estimating r_1 using either τ or T from time domain data and updated at each step of the simplex method. The inversion results using the constraint equation for r_1 as a function of τ_1, τ_2, τ_3, r_2, r_3, and τ are listed in Table III. The residual quantities, $\sqrt{\hat{E}/n}$, are less than 9 kHz. The values in parentheses are the differences with the expected values. The inverted thickness values are within -4 to 13 μm from the expected values for all samples. The inverted acoustic impedances are within -1.9% to 4.4% of the expected values.

For comparison purposes, thickness-only measurements (assuming known densities and velocities) by minimizing the difference square of the magnitude of the transfer function,

TABLE IV. Thickness measurement results using the Kinra *et al.* (Ref. 7) algorithm.

Property	304SS/Al/304SS	436SS/Al/304SS	Ag/Cu/Ni
h_1 (mm$\pm\mu$m)	0.387±1 (+12 μm)	0.434±2 (−4 μm)	0.314±9 (−30 μm)
h_2 (mm$\pm\mu$m)	1.523±1 (−7 μm)	1.828±2 (−5 μm)	1.144±9 (+12 μm)
h_3 (mm$\pm\mu$m)	0.363±1 (−12 μm)	0.384±1 (−4 μm)	0.306±3 (+21 μm)
h (mm$\pm\mu$m)	2.273 (−6 μm)	2.646 (−16 μm)	1.764 (−2 μm)

TABLE V. Inversion results for resonant frequencies measured in reflection.

Property	304SS/Al/304SS	436SS/Al/304SS	Ag/Cu/Ni
$\sqrt{\hat{E}/n}$ (KHz)	4.80	4.63	6.50
τ_1 (μs)	0.0661 (+1.0 ns)	0.0710 (−0.8 ns)	0.0930 (+0.03 ns)
τ_2 (μs)	0.2383 (−0.8 ns)	0.2864 (−0.01 ns)	0.2422 (−0.7 ns)
τ_3 (μs)	0.0648 (−0.3 ns)	0.0677 (+0.3 ns)	0.0507 (+0.08 ns)
r_2	0.382 (+0.63%)	0.372 (+1.13%)	1.101 (+3.14%)
r_3	2.620 (−0.54%)	2.687 (+2.00%)	1.209 (+1.20%)

performed by Kinra *et al.*,[22] are given in Table IV. The values with a "±" sign indicate precision and the values inside the parentheses indicate the difference with the expected values from Table I. The inverted thickness values are within $\pm 12\ \mu$m for 304SS/Al/304SS and 436SS/Al/304SS, and within $\pm 30\ \mu$m for the Ag/Cu/Ni sample.

The inversion results using the constraint equation for r_1 as a function of r_2, r_3, and T are listed in Table V. The inversion for the impedance ratios r_2 and r_3 are very close to the inverted results in Table III, but the inversion for r_1 is not. Furthermore the inverted thicknesses (assuming the densities are known) were one order less accurate. This can be easily understood from the fact that, using τ, r_1 is constrained in such a way that the total thickness remains accurate (see Eq. 2), which is not the case when r_1 is constrained using T.

As a note of caution on the interpretation of the inverted properties, the response in transmission is invariant to the direction of the normally-incident wave. Because of this, the resonant frequencies are also invariant to the forward or backward order of the layers. Therefore the error function has at least two equal minima where the parameters at these minima are related by

$$\begin{aligned} &\tau_1=\tau_3', \quad \tau_2=\tau_2', \quad \tau_3=\tau_1', \\ &r_1=r_3'r_2'r_1', \quad r_2=1/r_3', \quad r_3=1/r_2' \end{aligned} \tag{29}$$

This was observed for the 436SS/Al/304SS specimen, whose first and third layers have similar properties.

B. Inversion in reflection

First, the resonant frequencies are obtained from the phase of the transfer function in reflection. An example of a phase plot is shown in Fig. 6, for the 304SS/Al/304SS specimen using a 10 MHz transducer. The dots are the measured values and the solid line is a theoretical curve. The resonant frequencies were measured as the center frequency between the minimum point close to $-\pi$ and the maximum point close to $+\pi$.

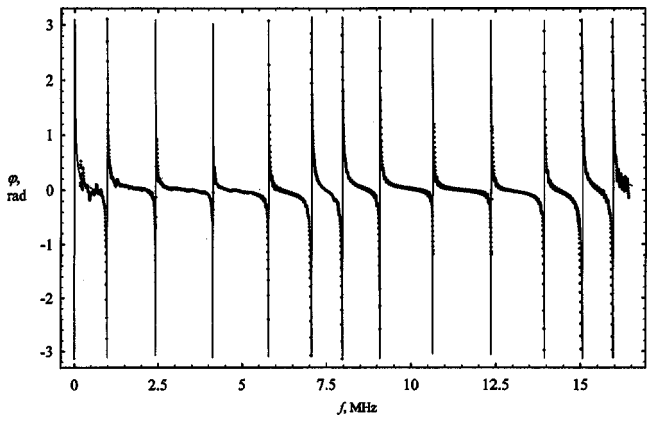


FIG. 6. Phase plot of the transfer function in reflection for 304SS/Al/304SS.

The inversion results for τ_1, τ_2, τ_3, r_2, and r_3 are given in Table V. The inversion results for 304SS/Al/304SS

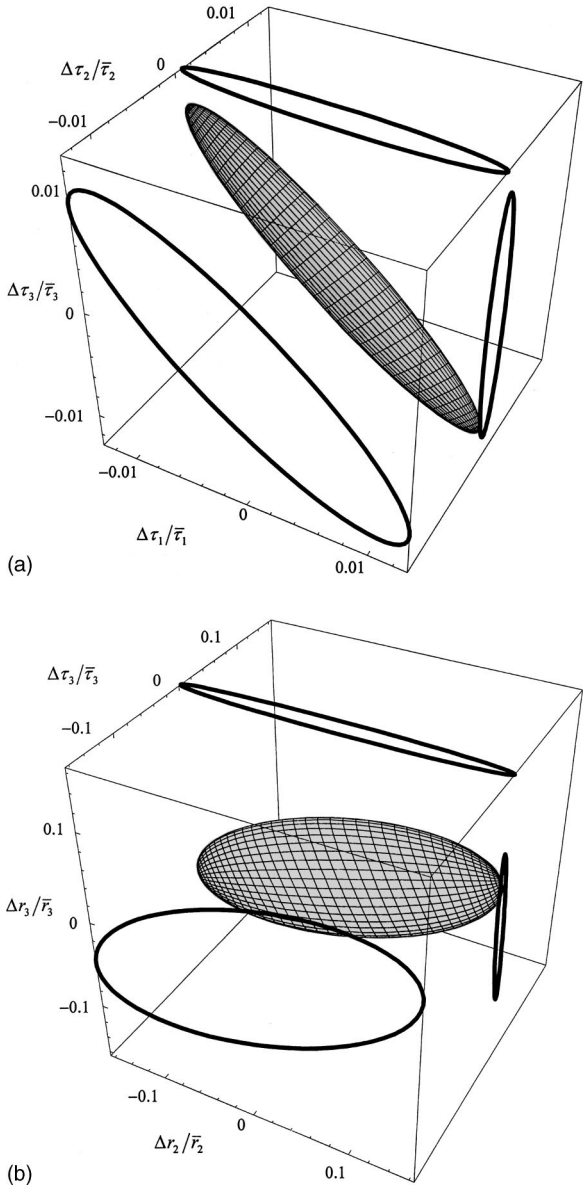


FIG. 7. (a) Error ellipses for the travel times using fluid-loaded resonances. (b) Error ellipses for r_2, r_3, and τ_3 using fluid-loaded resonances.

and 436/Al/304SS gave large errors for r_2 and r_3 of the order of $+10\%$. This was because the impedance difference and the thickness difference between the first and third layers is small and the resonant frequencies are insensitive to simultaneous changes of the impedance ratios r_2 and r_3 [see Fig. 4(b)]. For such a symmetric case the matrix $C^{-1}_{\alpha\beta}$ in Eq. (21) is close to being singular and the results of the inversion measurements are very sensitive to the measurement errors for the resonant frequencies. Therefore, for the 436SS/Al/304SS and 304SS/Al/304SS samples r_3 was constrained or set equal to $1/r_2$ in the inversion procedure. Consequently, for these samples, Table V shows the results for a four parameter inversion using the resonant frequencies subject to a constraint for r_3. For Ag/Cu/Ni, a five parameter inversion was carried out. For this sample all layers have different impedances and the converged results are as accurate as those obtained in transmission.

C. Errors of the inverted properties

An estimate of the errors of the inverted properties is obtained from the confidence ellipsoid given by Eq. (26). As an illustrative example consider the specimen 436SS/Al/304SS. For this specimen, the sensitivities of the fluid-loaded resonances to the travel times are about two orders of magnitude greater than those to the impedance ratios. Using $\sqrt{\hat{E}/n}=5.2$ kHz, the projections of a 5D ellipsoid onto 3D and 2D planes are obtained. A confidence value of $\gamma=0.95$ is used. The ellipses for the travel times are plotted in Fig. 7(a). The error bounds for the travel times are within ± 1 ns. The error bounds for τ_3, r_2, and r_3 are plotted in Fig. 7(b). The error bounds for the impedance ratios are of the order of 10%. These plots suggest that the true values for the travel times are within a nanosecond with 95% probability and that the true values for the impedance ratios are within about 10% with 95% probability. These estimates are in agreement with the actual deviations listed in Table III.

VI. CONCLUSIONS

A multiparameter inversion algorithm using the resonant frequencies of a layered medium was presented. Measurements for fluid-loaded resonances were carried out in the frequency range 1–20 MHz, that is at moderate frequencies where the smallest roundtrip travel time is of the order of the period of the highest frequency used. It was shown that not only travel times but also acoustic impedances can be simultaneously inverted using the resonant frequencies. Furthermore, from linearized error estimation, a criterion was obtained for the accuracy needed in the measurements of the resonant frequencies to invert for the desired acoustical properties at a given accuracy. The estimated error bounds obtained at a 95% probability were in agreement with the deviations for the inverted parameters from expected values. Inversion algorithm and error estimate is valid for any inversion problem with discrete number of measured quantities such as, for example, elastic moduli and internal friction measurements using normal mode resonance frequencies.[23]

ACKNOWLEDGMENTS

This paper is based on the work supported by the Texas Advanced Technology Program under Grant No. 9999-03-015 to Texas A&M University, College Station, TX., and National Science Foundation (NSF) Offshore Technology Research Center (OTRC). The authors wish to thank Yuquin Wang for providing the inversion results using the algorithm by Kinra *et al.*[9] and to Suraj Rawal for supplying the Ag/Cu/Ni sample.

[1] I. N. Komsky, J. D. Achenbach, G. Andrew *et al.*, ''An ultrasonic technique to detect corrosion in DC-9 wing box from concept to field application,'' Mater. Eval. **1995**, 848–852 (1995).

[2] S. E. Hanneman and V. K. Kinra, ''A new technique for ultrasonic nondestructive evaluation of adhesive joints: Part II. Experiment,'' Exp. Mech. **32**, 332–339 (1992).

[3] B. Tang, J. Mommaerts, and R. K. Duncan, ''Nondestructive evaluation of model adhesive joints by PVDF piezoelectric film sensors,'' Exp. Mech. **33**, 102–109 (1993).

[4] V. K. Kinra and C. Zhu, ''Ultrasonic nondestructive evaluation of thin (sub-wavelength) coatings,'' J. Acoust. Soc. Am. **93**, 2454–2467 (1993).

[5] P.-C. Xu, K.-E. Lindenschmidt, and S. A. Meguid, ''A new high-frequency analysis of coatings using leaky Lamb waves,'' J. Acoust. Soc. Am. **94**, 2954–2962 (1993).

[6] T. Kundu, ''Inversion of acoustic material signature of layered solids,'' J. Acoust. Soc. Am. **91**, 591–600 (1992).

[7] J. Krause, ''Thickness measurement on multilayered structures by SAW dispersion,'' Ultrasonics **32**, 195–199 (1994).

[8] O. Lenoir, J. L. Izbicki, P. Rembert, G. Maze, and J. Ripoche, ''Acoustic scattering from an immersed plane multilayer: Application to the inverse problem,'' J. Acoust. Soc. Am. **91**, 601–612 (1992).

[9] V. K. Kinra, P. T. Jaminet, C. Zhu, and V. R. Iyer, ''Simultaneous measurement of the acoustical properties of a thin-layered medium: The inverse problem,'' J. Acoust. Soc. Am. **95**, 3059–3074 (1994).

[10] O. I. Lobkis, A. Safaeinili, and D. E. Chimenti, ''Precision ultrasonic reflection studies in fluid-coupled plates,'' J. Acoust. Soc. Am. **99**, 2727–2736 (1996).

[11] V. K. Kinra and V. Dayal, ''A new technique for ultrasonic-nondestructive evaluation of thin specimens,'' Exp. Mech. **28**, 288–297 (1988).

[12] Clad Metals, Inc., 424 Morganza Rd., Canonsburg, PA 15317, USA.

[13] Polymetallurgical Corp., 262 Broad St., North Aftleboro, MA 02761, USA.

[14] V. K. Kinra and V. R. Iyer, ''Ultrasonic measurement of the thickness, phase velocity, density or attenuation of a thin viscoelastic plate. Part II: The inverse problem,'' Ultrasonics **33**, 111–122 (1995).

[15] D. Ensminger, *Ultrasonics: Fundamentals, Technology, Applications* (Marcel Dekker, New York, 1988).

[16] S. E. Hanneman and V. K. Kinra, ''A new technique for ultrasonic nondestructive evaluation of adhesive joints: Part I. Theory,'' Exp. Mech. **32**, 323–331 (1992).

[17] L. M. Brekhovskikh and O. A. Godin, *Waves in Layered Media 1*, (Springer-Verlag, Berlin, 1980), p. 32.

[18] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*: *The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1992).

[19] B. Ursin and K. A. Berteussen, ''Comparison of some inverse methods in wave propagation in layered media,'' Proc. IEEE **74**, 389–400 (1986).

[20] F. Habibi-Ashrafi and J. M. Mendel, ''Estimation of parameters in lossless layered media systems,'' IEEE Trans. Autom. Control **AC-27**, 31–48 (1982).

[21] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis* (Prentice-Hall, New Jersey, 1992), p. 298.

[22] V. K. Kinra, Y. Wang, C. Zhu, and S. P. Rawal, ''Simultaneous reconstruction of the acoustical properties of a layered medium: The inverse problem,'' in *Review of Progress in QNDE*, edited by D. O. Thompson and D. E. Chimenti 14, 1433–1440 (1995).

[23] Y. Sumino, I. Ohno, T. Goto, and M. Kumazawa, ''Measurement of elastic constants and internal frictions on single-crystal mgo by rectangular parallelepiped resonance,'' J. Phys. Earth **24**, 263–270 (1976).

Measurement of dependence of backscatter coefficient from cylinders on frequency and diameter using focused transducers—with applications in trabecular bone

Keith A. Wear[a)]
U.S. Food and Drug Administration, Center for Devices and Radiological Health, HFZ-142, 12720 Twinbrook Parkway, Rockville, Maryland 20852

(Received 25 April 2003; revised 3 October 2003; accepted 20 October 2003)

A theory for the elastic scattering response from a cylinder insonified by a plane wave was previously derived by Faran. In the present paper, the empirical relationship between Faran's theory and measurements of backscatter coefficient from cylindrical targets using focused transducers is investigated. Experimental measurements of dependence of backscatter coefficient on frequency and diameter for nylon wires are reported. It is found that, under certain conditions (including weak, incoherent scattering), backscatter coefficient measurements from collections of cylindrical scatterers may be meaningfully compared with Faran's model predictions. At low frequencies, the theory and experimental measurements exhibit similar dependences on frequency and diameter, provided that the scatterers are not too densely packed. At higher frequencies, the fine structure of Faran's predictions becomes difficult to reproduce experimentally with a focused transducer. Implications regarding applications to characterization of trabecular bone are discussed. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1631943]

PACS numbers: 43.20.Fn [FD] Pages: 66–72

I. INTRODUCTION

Measurements of scattering properties have been demonstrated to be useful for the characterization of many biological tissues (Chivers, 1977; Segal and Greenleaf, 1984; Shung, 1993). Many studies have involved measurement of the backscatter coefficient, which is defined as the differential scattering cross section per unit volume for a scattering angle of 180 degrees (Ishimaru, 1978; O'Donnell, 1981; Madsen, 1984).

Numerous algorithms for measurement of backscatter coefficient have been reported. These techniques often assume point-like scatters producing quasispherical scattered waves which may be detected efficiently with spherically focused transducers. Some techniques have been rigorously validated using Faran's theory of elastic scattering from spherical scatterers (Faran, 1951; O'Donnell, 1981; Madsen, 1984). Faran's theory predicts the magnitude of a spherical wave scattered by a sphere, or a cylindrical wave scattered by a cylinder, in response to an incident plane wave. (In the second case, the axis of the cylinder is assumed to be perpendicular to the wave propagation direction.) Faran's equations for spherical scatterers are consistent with the assumptions underlying many backscatter coefficient measurement methods. For example, Faran's assumption of plane-wave insonification is reasonable for small scatterers insonified by focused transducers provided that the variation of the amplitude and phase of the interrogating beam is negligible across the surface of the scatterer. This is often true when the scatterers are much smaller than the beam cross section. Thus, for small spherical scatterers, comparison of experimental measurements of backscatter coefficient to Faran's predictions is relatively straightforward.

The case of cylinders is somewhat more difficult than that for spheres. In Faran's model, cylinders have infinite extent. In practical experiments, the effective scatterer becomes that portion of the cylinder for which the transducer is most sensitive (i.e., the intersection of the beam cross section with the cylinder). But considerable variation of amplitude and phase of the incident beam is likely across the surface of the cylinder, violating Faran's premise of plane-wave insonification. Moreover, the differential scattering cross section (upon which the backscatter coefficient is based) is defined only under conditions in which a scattered field is measured sufficiently distant from the scatterer that it behaves as a spherical wave (Ishimaru, 1978). Since the effective scatterer spans the entire resolution volume, the distance required so that the scattered wave may be thought of as quasispherical is much greater for cylinders than for point-like scatterers. Therefore, in the case of interrogating a cylindrical scatterer at the focal region of a focused transducer, the scattered wave may not closely approximate a quasispherical form by the time it is received. Thus, the association of Faran's predictions with measured backscatter coefficients is not as clear as in the case of spherical scatterers.

The objective of this paper is to experimentally investigate the correspondence between Faran's theory and backscatter coefficient measurements using focused transducers and cylindrical targets. Only the dependences of backscatter coefficient on frequency and diameter are considered here, not the more challenging topic of absolute magnitude. This subject is particularly relevant for the characterization of trabecular bone. Previous measurements of frequency-dependent backscatter from human calcaneus have suggested that scatterers from trabecular bone behave more like cylin-

a)Electronic mail: kaw@cdrh.fda.gov

 0001-4966/2004/115(1)/66/7/$20.00

ders than spheres (Wear, 1999; 2000). Trabecular bone is known to contain rod-like trabeculae which could help explain this observation. Trabecular bone also contains plate-like structures. The relative preponderance of rod-like structures to plate-like structures increases with age.

Previous experiments have offered much validation for Faran's theory regarding measurements made with unfocused transducers. Faran himself reported measurements of scattering from metal cylinders at 1 MHz (Faran, 1951). A hydrophone was used to measure the angle dependence of scattering. Excellent agreement between theory and experiment was reported. Ninety-degree scattering from cylindrical (copper and nylon) targets was reported by Lee and coworkers (Lee, 1978), who also found good agreement between their measurements and Faran's theory.

Faran and Lee *et al.* used small-aperture transducers to insonify cylinders situated in the far field. This resulted in relatively uniform insonification over a substantial length of the cylinder, and, as far as their experiments were concerned, could be taken to approximate plane-wave insonification.

Faran's and Lee *et al.*'s pioneering work preceded the advent of algorithms for backscatter coefficient measurement using focused transducers in the early 1980's (O'Donnell, 1981; Lizzi, 1983; Campbell and Waag, 1983; Madsen, 1984; Ueda, 1985). Unlike the studies of Faran and Lee *et al.*, the present study is based on the measurement of backscatter coefficient using focused (rather than planar) transducers. These days, focused transducers are generally preferred for imaging applications and for measurements on small samples because of their superior lateral resolution. Also, in highly attenuating media, focused transducers are often required in order to overcome attenuation in tissue between the transducer and the region of interest.

Another distinction between the present study and that of Lee *et al.* is that Lee *et al.* examined scattering at 90 deg instead of backscatter. It will be seen that, in the range of frequencies investigated, backscatter from nylon scatterers exhibits a far more complicated form than the quasi-Gaussian response (backscattered intensity vs frequency) Lee *et al.* encountered at 90 deg, hence making the conformity between theory and experiment somewhat more challenging in the present study. (It should be noted however that Lee *et al.* did observe more complex behavior for copper wires). Finally, agreement between theory and experiment is investigated in the present study over a greater dynamic range of backscatter (two decades) than in Lee's work (approximately one decade).

Ueda and Akita reported a diffraction correction for backscatter coefficient using echoes scattered by a fine wire (Ueda, 1986). Their method showed excellent agreement with theoretical predictions. Their analysis was restricted to the case for which the diameter of the wire is much smaller than the wavelength of ultrasound. In the present study, the ratio of the cylinder diameter to the wavelength ranges from about 0.05 to 0.8.

There have been several other important contributions regarding scattering from cylinders that neither involve backscatter coefficient measurement nor offer comparisons to Faran's theory. Li and Ueda have provided a theoretical treatment of scattering of a spherical wave by an elastic circular cylinder (Li, 1990a). The same authors subsequently explored the measurement of scattering from cylinders using cylindrical shell transducers (Li, 1990b). Sheng and Hay investigated backscatter from thin wires situated in the transducer farfield, where the incident wave may be modeled as a spherical wave (Sheng, 1993). Raum and O'Brien developed a technique for determining the spatial and temporal transmit–receive field distributions of a spherically focused high-frequency transducer using a (tungsten) wire target (Raum, 1997).

II. METHODS

Backscatter coefficients are generally associated with media that contain multiple individual scatterers distributed throughout a volume. Measurement methods usually assume incoherent scattering (O'Donnell, 1981; Madsen, 1984). In other words, the phase differences between echoes emanating from scatterer pairs are assumed to be uniformly distributed from 0 to 2π. In this case, the expected received power is simply the sum of the powers scattered by individual scatterers situated within the resolution volume.

A natural approach for investigating this problem experimentally might be to acquire data from phantoms consisting of multiple randomly positioned cylinders per resolution volume. However, design and construction of such a phantom would represent a substantial challenge. In the present study, an alternative but essentially equivalent approach was adopted for expediency. Signals were acquired from single cylinders placed at the focus of the transducer. Under the assumptions of weak, incoherent scattering, the dependence of backscatter on cylinder *diameter* would be expected to be the same for a single scatterer as for a volume containing randomly distributed scatterers. However, the dependence of backscatter on *frequency* would be expected to be different in the two cases since the size of the resolution volume is frequency dependent, and therefore the effective number of scatterers contained within the resolution volume is frequency dependent. In order to account for this effect, measured signal power spectra may be multiplied by a presumed frequency-dependent number of scatterers per resolution cell (see below for details). This approach is consistent with the underlying assumptions for backscatter coefficient measurement and avoids the arduous task of designing and building a phantom containing randomly positioned cylinders. This approach also assumes that the system point-spread function is shift invariant throughout the measurement volume. Backscatter coefficients compensated in this way are referred to below as ''expected backscatter coefficients.''

Nylon fishing lines of various diameters (nominally 0.006, 0.008, 0.010, 0.011, 0.012, 0.014, 0.016, and 0.018 in.) were interrogated in a water tank in pulse–echo mode. Actual diameters were measured using calipers. Each line had a fishing weight attached to the bottom so that it could be hung vertically. The ultrasound beam propagation direction was perpendicular to the nylon line. In order to compute the predicted backscatter according to Faran's theory, assumed values for Poisson's ratio, density (ρ), and longitudinal velocity (c_l) for the scattering target were required. Pois-

TABLE I. Properties of Panametrics ultrasonic transducers used.

Model number	Center frequency (MHz)	Diameter (inches)	Focal length (inches)
V391	0.5	1.125	2.1
V302	1.0	1.0	2.0
V305	2.25	0.75	1.5

son's ratio for nylon was assumed to be 0.39 and c_l was taken to be 2600 m/s (Ondacorp, 2003). Density (ρ) was obtained from measured masses (m) of the spools of fishing line. The measured mass of each spool of fishing line was the sum of the mass of the plastic spool (m_s) plus that of the nylon line (ρv) wound around it: $m = m_s + \rho v$, where the nylon wire volume v was $\pi r^2 h$, r was the nylon line radius, and h was the length. Since r (measured with calipers) and h (specified by the manufacturer) were known, m_s and ρ could be estimated from a linear regression to measurements of m and v from the eight spools. The density estimated in this way was 1.3 g/cc. The square of the correlation coefficient for the linear fit was 0.98 (95% confidence interval: 0.96–1.00), indicating a high uniformity of densities among the eight nylon lines. The theoretical predictions for scattering from a cylindrical nylon target was generated numerically (MATLAB) using Faran's theory (Faran, 1951) and these material values.

Three different Panametrics (Waltham, MA) ultrasound transducers, with center frequencies ranging from 500 kHz to 2.25 MHz, were used. See Table I. For low-frequency measurements (500 kHz), a Panametrics model 5077 PR pulser/receiver was used. For higher frequencies, a Panametrics model 5800 PR was used. Received ultrasound signals were digitized (8 bit, 10 MHz) using a LeCroy (Chestnut Ridge, NY) 9310C Dual 400-MHz oscilloscope and stored on computer (via GPIB) for off-line analysis. In order to ensure consistent alignment of nylon lines, the transducer orientation was always adjusted so as to maximize the amplitude of the received time-domain signal.

Backscatter coefficients were measured using a reference phantom method (Zagzebski *et al.*, 1993) as described previously (Wear, 1999). This method assumes that the location, size, and shape of the region of interest (defined by the gate duration in the axial dimension and the beamwidth in the lateral dimension) are identical for the specimen for which backscatter coefficient is to be measured, and the reference phantom (for which the backscatter coefficient is presumed to be known *a priori*). In this case, the effects of frequency-dependent diffraction are the same for both measurements and therefore cancel out when the ratio of power spectra is taken. A gate length of 1 cm was used for the reference phantom measurements.

As mentioned above, for the investigation of the expected dependence of backscatter coefficient on frequency, received power spectra from the nylon cylinders were multiplied by a frequency-dependent factor in order to simulate the situation in which numerous cylinders exist within a resolution cell. It is assumed here that all the hypothetical cylinders are oriented parallel to each other and perpendicular to the ultrasound propagation direction. Since the beamwidth is frequency dependent, the resolution cell size is frequency dependent. The effective number of cylinders per resolution cell N is given by

$$N = n_x n_z, \tag{1}$$

where n_x = effective number of scatterers contained across the lateral dimension of the resolution cell and n_z = effective number of scatterers contained across the depth dimension of the resolution cell. The latter quantity is given by

$$n_z = 1 + \Delta z/s, \tag{2}$$

where Δz is the length of the rectangular range gate, $\Delta z = c\tau/2$, c is the speed of sound, τ is the duration of the range gate, and s is the mean scatterer spacing. This factor (n_z) does not depend on frequency.

An expression for n_x may be obtained as follows. The intensity of the Franhofer diffraction pattern for a circular aperture (the so-called "Airy pattern") is given by (Goodman, 1968)

$$I(x) = \left(\frac{kl^2}{8z}\right)^2 \left[2\frac{J_1(klx/2z)}{klx/2z}\right]^2, \tag{3}$$

where J_1 is a Bessel function of the first kind, order one, $k = 2\pi/\lambda$, λ is the wavelength, l is the aperture diameter, x is the lateral dimension, and z is the distance from the aperture to the scatterer. This same intensity distribution has been derived for acoustic radiation. [See Kinsler *et al.*, 1982, Eq. (8.35), and substitute $a = l/2$, $\sin\theta = x/z$]. The effect of focusing is to bring the Franhofer pattern from the far field to the focal plane of the transducer. From the principle of reciprocity, the overall system sensitivity, as a function of scatterer position, is the square of the focal plane intensity distribution

$$I^2(x) = \left(\frac{kl^2}{8z}\right)^4 \left[2\frac{J_1(klx/2z)}{klx/2z}\right]^4. \tag{4}$$

For the purpose of estimating n_x, it was assumed that the effective number of scatterers across the lateral direction could be taken to be equal to 1 (the wire at the focal point) plus the expected number of scatterers to the left plus the expected number of scatterers to the right, given by

$$n_x = 1 + 2\Delta x/s, \tag{5}$$

where the half width half maximum (HWHM) of $I^2(x)$ is denoted by $\Delta x \approx 0.37zc/(lf)$ and f is frequency.

Two limitations to this approach should be acknowledged. First, it does not explicitly account for the fact that cylinders near the periphery of the beam are insonified with lower intensities than centrally positioned cylinders. However, it has previously been shown that, for the purpose of measuring backscatter coefficients, the effective number of scatterers may be accurately approximated simply by counting those scatterers included within the 3-dB beam cross section [O'Donnell and Miller, 1981, Eq. (15)]. This approximation is especially valid in the present application as the parameter of interest is only the *frequency dependence* (rather than the *absolute magnitude*) of the scatterer number. A second limitation to this approach is that it does not account for the fact that cylinders near the periphery of the

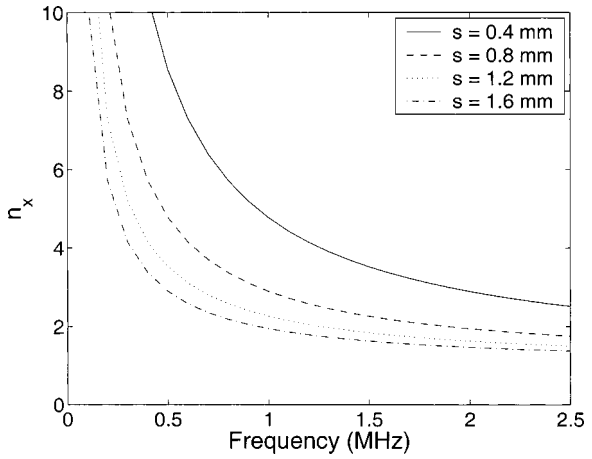


FIG. 1. The expected number of scatterers (n_x) across the lateral dimension as a function of frequency for four different assumed mean scatterer spacings (s).

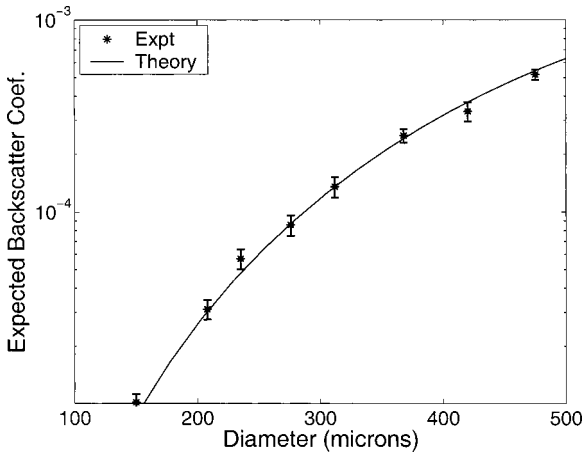


FIG. 2. Backscattered intensity at 500 kHz as a function of diameter of nylon line.

beam will have diminished effective heights (i.e., intersections between beam cross section and cylinder) and therefore potentially altered frequency-dependent backscattering properties. Again, it is assumed here that, for the purpose of estimating *frequency dependence* of scatterer number, the present approach is adequate. An experimental test of this assumption is given in the next section.

In Fig. 1, n_x is plotted as a function of frequency for four choices of assumed mean scatterer spacing for an $f2$ transducer and an assumed sound speed of 1480 m/s (water). It can be seen that for mean scatterer spacings greater than or equal to about 0.8 mm, the approximate mean trabecular spacing for human calcaneus (Ulrich, 1999), and frequencies greater than 1 MHz, the frequency dependence of n_x is relatively gradual (especially when compared to Faran's frequency response from a single cylinder, as will be seen in the next section). Under these conditions, any conclusions regarding the frequency dependence of backscatter coefficient will not depend critically on the assumed value for mean scatterer spacing.

III. RESULTS

Figure 2 shows experimental measurements of backscatter coefficient at 500 kHz as a function of diameter of nylon line. The solid line shows the theoretical variation predicted from Faran's theory (Faran, 1951). The theoretical curve was multiplied by a single diameter-independent scale factor in order to optimally fit the data. Therefore, only the extent of agreement between theory and experiment regarding *dependence of backscatter coefficient on scatterer diameter* may be assessed from Fig. 2. No meaningful conclusions may be drawn regarding the *absolute magnitude* of backscatter coefficient. Theory and experiment exhibit similar dependences on diameter.

Figure 3 shows expected frequency-dependent backscatter coefficients from the eight nylon wires (diameters shown in upper left of each panel). Also shown are the theoretical backscattered intensities computed using Faran's theory (solid line). Each theoretical curve was multiplied by a frequency-independent scale factor in order to fit the data. Therefore, only the extent of agreement between theory and experiment regarding *dependence of backscatter coefficient on frequency* may be assessed from Fig. 3. No meaningful conclusions may be drawn regarding the *absolute magnitude* of backscatter coefficient.

As described in the Methods section, experimental results (but not the theory) shown in Fig. 3 were scaled by the effective number of scatterers per resolution cell, assuming a particular value for scatterer spacing, 0.8 mm, which corresponds to the measurement reported by Ulrich *et al.* (1999) based on their microcomputed tomographic analysis of human calcaneus samples. (They reported a mean trabecular separation of 684 microns and a mean trabecular thickness of 127 microns. Mean scatterer spacing is the sum of these two values.) Good agreement may be seen in Fig. 3 between theory and experiment, especially for the thinner diameters and lower frequencies. The experimental measurements do not always exhibit sufficient resolution to faithfully reproduce Faran's theory. In particular, for the three thickest diameters, the fine structure of Faran's predictions at frequencies above about 1.5 MHz becomes noticeably difficult to reproduce experimentally.

In Fig. 4, the dependence of backscatter coefficient for the 150-micron wire as a function of frequency is shown in the low-frequency range. This diameter is just slightly larger than the mean value for trabecular thickness in human calcaneus reported by Ulrich *et al.* (1999) of 127 microns. Again, experimental results were scaled by the effective number of scatterers per resolution cell assuming a scatterer spacing of 0.8 mm (see Fig. 1). Also shown is a power-law fit to the data, backscatter coefficient$=Af^m$, where $m=2.84$. This approximate cubic dependence is similar to theoretical predictions by Lizzi *et al.* (1996) for thin quasicylindrical scatterers.

IV. DISCUSSION: APPLICATIONS IN BONE

In a previous study (Wear, 1999), backscatter coefficients were measured from human calcaneus *in vitro*. These frequency-dependent backscatter coefficients were compared to Faran's predictions. This analysis was based on the plausible but unproven assumption that the two frequency-

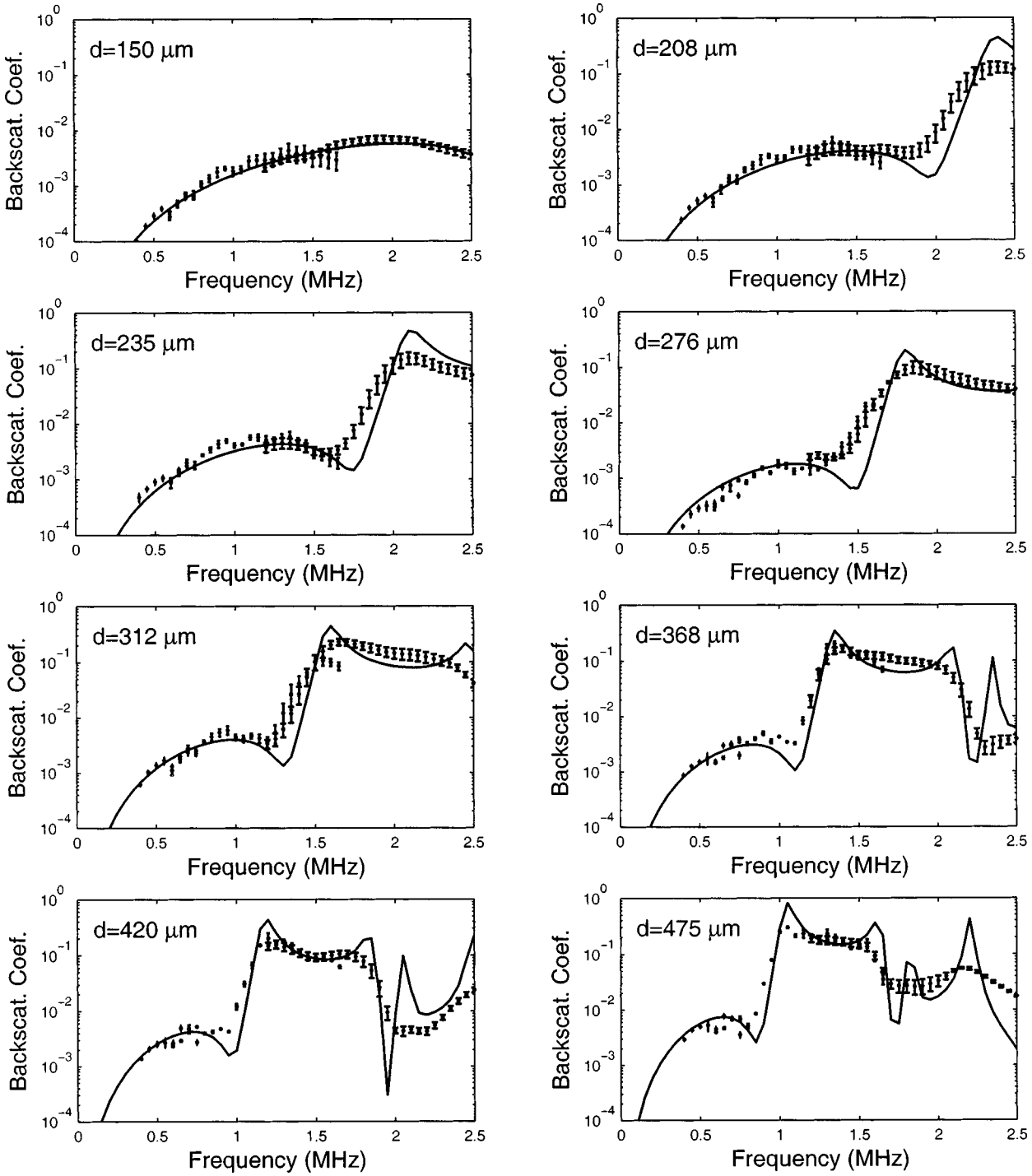


FIG. 3. Backscattered intensity from nylon wires (diameter shown in upper left of each panel). Also shown are the theoretical backscattered intensities computed using Faran's theory (solid line).

dependent functions should be roughly proportional to each other. The present study offers strong empirical evidence to support this assumption. The present study suggests that backscatter coefficient measurements exhibit similar dependences on frequency and diameter as Faran's model predicts, provided that the scatterers are not too densely packed. Although the present study neglected attenuation that may arise from a hypothetical ensemble of cylinders throughout the range gate, its results may be extended to highly attenuating media such as bone because backscatter coefficient measurements from tissues may be accurately compensated for the effects of attenuation [O'Donnell and Miller, 1981, Eq. (17)].

The reference phantom method for measurement of backscatter coefficient used in this study has been experimentally validated using homogeneous phantoms containing spherical or quasispherical scatterers (Zagzebski, 1993). The present study adds evidence that it is also appropriate for media containing cylindrical scatterers.

Faran's model predicts that, at low frequencies, backscattered intensity is approximately proportional to fre-

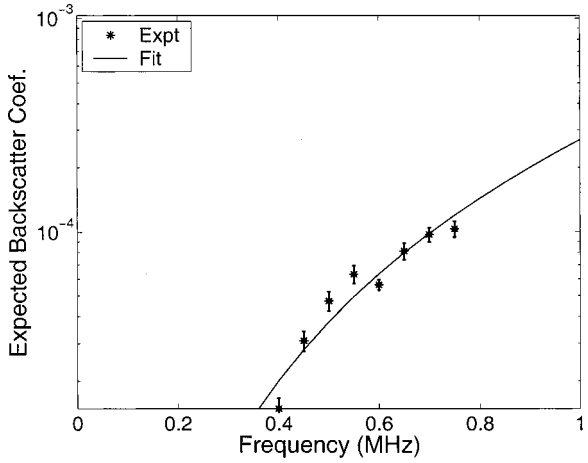


FIG. 4. The dependence of expected backscatter coefficient for the 150-micron (0.006 in.) wire as a function of frequency. Also shown is a power-law fit, backscatter coefficient $= Af^m$, where $m = 2.84$.

quency cubed. This effect is seen in the present study in Fig. 4, which shows the low-frequency backscatter coefficient for the 150-micron-diameter nylon wire. In this case, the exponent of the power-law fit was found to be 2.84 (approximately cubic). This value is similar to that reported in human calcaneus, 3.26 ± 0.20 (Wear, 1999). Although trabeculae in human calcaneus are oriented in a variety of directions, they are all approximately perpendicular to the ultrasound propagation direction provided that bones are interrogated in the medio-lateral orientation (as is the case in the experiments above and in most commercial bone sonometers).

Empirical power-law exponents for human calcaneus in the low-frequency range are a little higher than the approximately cubic variation predicted by the cylinder model -3.26 (Wear, 1999) and 3.38 (Chaffai *et al.*, 2000). The difference of 0.3–0.4 may be due in part to multiple scattering, which would have the effect of applying additional high-pass (frequency-cubed) filtering to echoes (via multiple scatters), resulting in measured backscatter coefficients with more rapid than cubic frequency dependence (Wear, 1999). This effect would be expected to be rather small because multiply scattered waves tend to traverse longer path lengths than singly scattered waves, and trabecular bone has a very high attenuation coefficient. Another potential contribution to the difference between theory and measurements results from trabecular bone having a more complex geometry than the cylinder model would suggest. Rod-like structures within trabecular bone exhibit a wide range of lengths and in some cases will be shorter than a beamwidth. Moreover, trabeculae are not perfectly straight (as assumed in the model) but are somewhat curved and jagged. Furthermore, in addition to rod-like structures, trabecular bone also contains plate-like components. Nevertheless, cylinder-like objects (trabeculae) are responsible for an appreciable fraction of the scattering. In addition, only those plate-like structures oriented approximately perpendicular to the ultrasound propagation direction can measurably affect the frequency dependence of scattering. Finally, plates are comparatively rare in bones from older subjects, upon which the data from the studies mentioned above are based.

Trabecular thickness is an important determinant of osteoporotic fracture risk. Noninvasive assessment of trabecular thickness could potentially yield useful diagnostic information. A recent study has shown that the dependence of backscatter coefficient measurements from human calcaneus samples on trabecular thickness is similar to that predicted using Faran's model (Wear, 2003). The theory predicts that, in the range of morphological and material properties expected for trabecular bone, the backscatter coefficient at 500 kHz should be approximately proportional to trabecular thickness to the power of 2.9. A power-law fit to empirical data suggested that backscatter coefficient measurements (from 43 human calcaneus samples) varied as trabecular thickness to the 2.8 power (95% confidence interval: 1.7–3.9).

V. CONCLUSION

Under certain conditions (including weak, incoherent scattering), backscatter coefficient measurements from collections of cylindrical scatterers may be meaningfully compared with Faran's model predictions of scattered wave intensities resulting from plane-wave insonification. At low frequencies, the two functions exhibit similar dependences on frequency and diameter, provided that the scatterers are not too densely packed. At higher frequencies, the fine structure of Faran's predictions becomes difficult to reproduce experimentally.

ACKNOWLEDGMENTS

The author is grateful for funding provided by the US Food and Drug Administration Office of Women's Health. The author thanks Timothy Hall, University of Kansas, for providing the reference phantom used in this study.

Campbell, J. A., and Waag, R. C. (**1983**). "Normalization of ultrasonic scattering measurements to obtain average differential scattering cross sections for tissues," J. Acoust. Soc. Am. **74**, 393–399.

Chaffai, S., Roberjot, V., Peyrin, F., Berger, G., and Laugier, P. (**2000**). "Frequency dependence of ultrasonic backscattering in cancellous bone: Autocorrelation model and experimental results," J. Acoust. Soc. Am. **108**, 2403–2411.

Chivers, R. C. (**1977**). "The scattering of ultrasound by human tissues—some theoretical models," Ultrasound Med. Biol. **3**, 1–13, Appendix.

Faran, J. J. (**1951**). "Sound scattering by solid cylinders and spheres," J. Acoust. Soc. Am. **23**, 405–418.

Goodman, J. W. (**1968**). *Introduction to Fourier Optics* (McGraw-Hill, New York).

Ishimaru, A. (**1978**). *Wave Propagation and Scattering in Random Media* (Academic, New York), Chap. 2.

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**1982**). *Fundamentals of Acoustics* (Wiley, New York), Chap. 8, p. 179.

Lee, P. P. K., Waag, R. C., and Hunter, L. P. (**1978**). "Swept-frequency diffraction of ultrasound by cylinders and arrays," J. Acoust. Soc. Am. **63**, 600–606.

Li, T., and Ueda, M. (**1990a**). "Sound scattering of a spherical wave incident on a cylinder," J. Acoust. Soc. Am. **87**, 1871–1879.

Li, T., and Ueda, M. (**1990b**). "Analysis of echoes from a cylinder that includes the directivity of the transmitter and receiver," J. Acoust. Soc. Am. **87**, 1880–1884.

Lizzi, F. L. Greenebaum, M., Feleppa, E. J., Elbaum, M., and Coleman, D. J. (**1983**). "Theoretical framework for spectrum analysis in ultrasonic tissue characterization," J. Acoust. Soc. Am. **73**, 1366–1373.

Lizzi, F. L., Astor, M., Kalisz, A., Liu, T., Coleman, D. J., Silverman, R., Ursea, R., and Ro, M. (**1996**). "Ultrasonic spectrum analysis for assays of

different scatter morphologies: Theory and very-high frequency clinical results,'' Proceedings of the 1996 IEEE Ultrasonics Symposium, Vol. 2, pp. 1155–1159.
Madsen, E. L., Insana, M. F., and Zagzebski, J. A. (**1984**). ''Method of data reduction for accurate determination of acoustic backscatter coefficients,'' J. Acoust. Soc. Am. **76**, 913–923.
O'Donnell, M., and Miller, J. G. (**1981**). ''Quantitative broadband ultrasonic backscatter: An approach to nondestructive evaluation in acoustically inhomogeneous materials,'' J. Appl. Phys. **52**, 1056–1065.
Ondacorp, http://ondacorp.com/tables/plastics.htm
Raum, K., and O'Brien, Jr., W. D. (**1997**). ''Pulse–echo field distribution measurement technique for high-frequency ultrasound sources,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 810–815.
Sehgal, C. M., and Greenleaf, J. F. (**1984**). ''Scattering of ultrasound by tissues,'' Ultrason. Imaging **6**, 60–80.
Sehgal, C. M. (**1993**). ''Quantitative relationship between tissue composition and scattering of ultrasound,'' J. Acoust. Soc. Am. **94**, 1944–1952.
Sheng, J., and Hay, A. (**1993**). ''Spherical wave backscatter from straight cylinders: Thin-wire standard targets,'' J. Acoust. Soc. Am. **94**, 2756–2765.
Shung, K. K., and Thieme, G. A. editors (**1993**). *Ultrasonic Scattering in Biological Tissues* (CRC Press, Boca Raton, FL).
Ueda, M., and Akita, M. (**1986**). ''Diffraction correction for backscattering coefficient measurement using echoes scattered by a fine wire,'' Proc. 1986 IEEE Ultrasonics Symp. 0090-5607/86/0000-0913.
Ueda, M., and Ozawa, Y. (**1985**). ''Spectral analysis of echoes for backscattering coefficient measurement,'' J. Acoust. Soc. Am. **77**, 38–47.
Ulrich, D., van Rietbergen, B., Laib, A., and Ruegsegger, P. (**1999**). ''The ability of three-dimensional structural indices to reflect mechanical aspects of trabecular bone,'' Bone **25**, 55–60.
Wear, K. A. (**1999**). ''Frequency dependence of ultrasonic backscatter from human trabecular bone: theory and experiment,'' J. Acoust. Soc. Am. **106**, 3659–3664.
Wear, K. A. (**2000**). ''Anisotropy of ultrasonic backscatter and attenuation from human calcaneus: Implications for relative roles of absorption and scattering in determining attenuation,'' J. Acoust. Soc. Am. **107**, 3474–3479.
Wear, K. A., and Laib, A. (**2003**). ''The relationship between ultrasonic scattering and microarchitecture in human calcaneus,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control (to be published).
Zagzebski, J. A., Yao, L. X., Boote, E. J., and Lu, Z. F. (**1993**). ''Quantitative Backscatter Imaging,'' in *Ultrasonic Scattering in Biological Tissues*, edited by K. K. Shung and G. A. Thieme (CRC Press, Boca Raton, FL).

The role of the coupling term in transient elastography

Laurent Sandrin[a)]
Echosens, Research & Development Department, 42 rue Monge, 75005 Paris, France

Didier Cassereau[b)] and Mathias Fink[c)]
Laboratoire Ondes et Acoustique, Université Paris VII/ESPCI, CNRS UMR 7587 10 rue Vauquelin, 75231 Paris Cedex 05, France

(Received 19 September 2002; revised 11 July 2003; accepted 30 October 2003)

The transient radiation of low-frequency elastic waves through isotropic and homogeneous soft media is investigated using the Green's function approach. A careful analysis of the coupling term is performed and yields the introduction of a very near field region in which its amplitude behaves as $1/r$. To address the calculation of impulse responses, a simplified Green's function is proposed for semi-infinite media and compared to exact solutions. Impulse response calculations are successfully compared with experimental measurements obtained for circular radiators of different diameters using transient elastography. Results presented in this paper provide a better understanding of the role of the coupling term in elastography and should be used to compensate diffraction and coupling effects observed in transient elastography. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1635412]

PACS numbers: 43.20.Gp, 43.20.Jr, 43.20.Px [YHB] Pages: 73–83

I. INTRODUCTION

Low-frequency shear waves have led to the development of promising tools in tissue characterization. The interest in the propagation of shear waves in soft tissues lies in the relation between the shear wave velocity and the Young's modulus more generally referred to as elasticity. It is now well known that the pathological state of soft tissues is strongly correlated with changes in stiffness. Dynamic elastographic techniques are based on the investigation of low-frequency shear waves in soft tissues. Shear waves are usually sent using mechanical vibrators placed at the surface of the body. These vibrators are excited either monochromatically as in sonoelastography[1–4] and magnetic resonance elastography (MRE)[5,6] or transiently as in ultrasound-based transient elastography.[7–11]

Transient elastography has been developed since 1994. Catheline *et al.*'s[7,8] first setup includes an ultrasonic transducer placed in front of a circular mechanical vibrator. The mechanical vibrator is used to send a low-frequency (typically 50 Hz) transient elastic wave while the ultrasonic transducer acquires rf lines at a repetition frequency of 1000 Hz. The displacements induced in the medium by the low-frequency elastic wave are obtained by comparison of the successive rf lines. A longitudinally polarized elastic wave whose velocity varies with elasticity is observed. This system proved efficiencent to estimate the elasticity of soft homogeneous and isotropic media. It was further improved by associating the transducer and the vibrator into a single element as explained by Sandrin *et al.*[10] Furthermore, transient elastography was applied to 2D elasticity estimation after the development of an ultrafast ultrasonic imaging scanner.[11]

Little information is found on the theoretical background of low-frequency elastic waves in soft media. However, many techniques[12–18] have been developed in order to calculate the transient ultrasonic fields radiated by arbitrary surface sources in solids. Their development is mainly due to ultrasonic field study for medical applications, nondestructive evaluation, and seismology. These techniques can be adapted to the case investigated in the present paper by changing the frequency of the excitation signal and the values of compressional and shear waves velocities in the medium. Indeed in traditional solids, shear wave velocity is of the same order as the compressional wave velocity while in soft tissues the shear wave velocity is much smaller than the compressional wave velocity.

In the present paper, the Green's function in soft media is evaluated based on the Green's function in homogeneous, isotropic and infinite solids.[19–21] Since only soft media are studied, conclusions may be irrelevant for classical solids. The role of the coupling term and its singular amplitude behavior in the very near field are investigated. We show that the Green's function associated with an infinite medium may be applied to the case of semi-infinite medium. Theoretical results are in very good agreement with experimental measurements obtained in tissue-mimicking phantoms. The results point out the role of diffraction and coupling effects that may induce biases in elasticity estimation performed with dynamic elastographic techniques.

In the second section of this paper, an analytical Green's function in infinite homogeneous and isotropic solids is described. A physical interpretation and analysis of the coupling term are proposed. The case of soft media is investigated. Calculations show that the coupling term behaves as r^{-1} in a very near field region and as r^{-2} in the classical near field. The third section deals with the computation of the impulse response and yields an analytical expression of the impulse response on the axis of a circular piston vibrator. In the fourth section, the use of a similar expression of the

[a)]Electronic mail: laurent.sandrin@echosens.com
[b)]Electronic mail: didier.cassereau@loa.espci.fr
[c)]Electronic mail: mathias.fink@espci.fr

infinite medium Green's function is proposed as a simplified model adapted to the case of a semi-infinite medium. Green's function obtained with the simplified model are compared to exact Green's function obtained by Gakenheimer[18] on the axis of a point load source. In the last section the theoretical impulse responses based on the simplified model are compared to experimental impulse responses obtained with transient elastography.

II. GREEN'S FUNCTION IN AN INFINITE MEDIUM

In this section, the Green's function associated with a point source in an infinite isotropic, homogeneous medium is introduced. A careful analysis of the near field coupling term of the Green's function in an infinite medium is proposed. After a general description of the coupling term, the analysis focuses on the case of soft media and details the particular behavior of the near field coupling term.

A. Near field coupling term

The Green's function associated with a point source links the *i*th component of the displacement vector to the stress applied in the *j* direction at point *O*:

$$G_{ij}(\mathbf{r},t)=G_{ij}^{c}(\mathbf{r},t)+G_{ij}^{s}(\mathbf{r},t)+G_{ij}^{cs}(\mathbf{r},t). \tag{1}$$

The Green's function in a homogeneous, isotropic and infinite solid initially introduced by Love,[19] Achenbach,[20] and Aki and Richards.[21] It is composed of three terms: the compressional term G^c, the shear term G^s, and the coupling term G^{cs}:

$$\begin{aligned}
G_{ij}^{c}(\mathbf{r},t)&=\frac{1}{4\pi\rho V_c^2}\gamma_i\gamma_j\frac{1}{r}\delta\left(t-\frac{r}{V_c}\right),\\
G_{ij}^{s}(\mathbf{r},t)&=-\frac{1}{4\pi\rho V_s^2}(\gamma_i\gamma_j-\delta_{ij})\frac{1}{r}\delta\left(t-\frac{r}{V_s}\right),\\
G_{ij}^{cs}(\mathbf{r},t)&=\frac{1}{4\pi\rho}(3\gamma_i\gamma_j-\delta_{ij})\frac{1}{r^3}\int_{r/V_c}^{r/V_s}\tau\delta(t-\tau)d\tau,
\end{aligned} \tag{2}$$

where $r=|\mathbf{r}|$, $\gamma_i=\partial r/\partial i$ ($i=x$, y or z), δ_{ij} is the Kronecker symbol ($\delta_{ij}=1$ when $i=j$), and ρ is the mass density. The compressional term G^c and the shear term G^s are classical far field terms, respectively purely longitudinal and purely transverse. They behave as r^{-1}. Their contributions occur respectively at times $t_c=r/V_c$ and $t_s=r/V_s$. The behavior of the coupling term is more complex because of the ramp that extends between times t_c and t_s. It will be shown that the r^{-3} coefficient does not result in a r^{-3} decay. The coupling term is neither irrotational nor solenoidal.

A few transformations of the coupling term G^{cs} in Eq. (2) are made in order to study its behavior. In a first step, the transformation $r'=\tau V_s$ is performed:

$$\begin{aligned}
G_{ij}^{cs}(\mathbf{r},t)=&\frac{1}{4\pi\rho V_s^2}(3\gamma_i\gamma_j-\delta_{ij})\frac{1}{r^3}\\
&\times\int_{rV_s/V_c}^{r}r'\delta\left(t-\frac{r'}{V_s}\right)dr'.
\end{aligned} \tag{3}$$

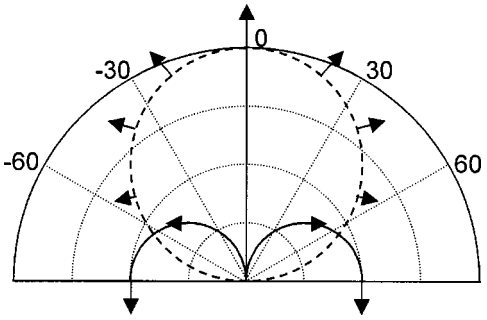


FIG. 1. Polar diagrams of the longitudinal (dotted line) and tangential (plain line) components of the near field coupling term for an excitation parallel to the vertical axis.

This expression of the coupling term demonstrates that it is inversely proportional to the shear modulus $\mu=\rho V_s^2$ and is proportional to r^{-3}. The coupling term expression simplifies to

$$G_{ij}^{cs}(\mathbf{r},t)=\frac{1}{4\pi\rho V_s^2}\times\delta_{ij}(\mathbf{r})\times f(r,t), \tag{4}$$

where $d_{ij}(\mathbf{r})$ is the directivity coefficient. $d_{ij}(\mathbf{r})$ and $f(r,t)$ satisfy the following equations:

$$d_{ij}(\mathbf{r})=3\gamma_i\gamma_j-\delta_{ij}, \tag{5}$$

$$f(r,t)=\frac{1}{r^3}\int_{rV_s/V_c}^{r}r'\delta\left(t-\frac{r'}{V_s}\right)dr'. \tag{6}$$

The directivity coefficient neither depends on the medium elastic properties nor on the distance r. The directivity patterns of the longitudinal and tangential components of $d_{ij}(\mathbf{r})$ for an excitation along z are given in Fig. 1. A purely longitudinal term is observed in the direction of the excitation and a purely tangential term is observed in the direction perpendicular to the direction of the excitation. In other directions, the near field coupling term is neither longitudinal nor transverse. It must be noticed that the directivity patterns presented in Fig. 1 only represent the near field term which can be neglected in the far field compared to the compressional and shear terms. Concerning the second coefficient, f, the transformation $x'=r'/r$ in Eq. (6) gives

$$f(r,t)=\frac{1}{r}\int_{V_s/V_c}^{1}x'\delta\left(t-\frac{rx'}{V_s}\right)dx'. \tag{7}$$

Therefore, $f(r,t)$ is a temporal ramp weighted with an r^{-1} coefficient. In this new expression, one notices that the coupling term is proportional to r^{-1} which is in contradiction with Eq. (3). This emphasizes the confusion made by some authors who tried to deduce the behavior of the coupling term directly from Eq. (3). Indeed the temporal shape of the near field coupling term is not a Dirac distribution and the integral part must be carefully analyzed to deduce the behavior of the shear term. The integral part of Eq. (7) is a temporal ramp. The calculation of its surface $S(r)$ yields to

$$S(r)=\int_{-\infty}^{+\infty}\int_{V_s/V_c}^{1} x'\,\delta\left(t-\frac{rx'}{V_s}\right)dx'\,dt=\int_{V_s/V_c}^{1} x'\,dx'$$
$$=\frac{V_c^2-V_s^2}{2V_c^2}. \tag{8}$$

The noticeable point is that the surface $S(r)$ only depends on the elastic parameters of the medium. It remains constant whatever the distance r. It varies between $\frac{3}{8}$ in classical solids ($V_s \simeq V_c/2$) and $\frac{1}{2}$ in soft media ($V_s \ll V_c$).

B. Soft media case

The analysis is now restricted to soft media ($\mu \ll \lambda$). In this case the compressional term that behaves as λ^{-1} is negligeable versus the shear and coupling terms that behave as μ^{-1}. Equation (1) simplifies to

$$G_{ij}(\mathbf{r},t)\simeq G_{ij}^{s}(\mathbf{r},t)+G_{ij}^{cs}(\mathbf{r},t). \tag{9}$$

In this particular case, the slope of the ramp is $(V_s/r)^2$, and its duration and height are respectively r/V_s and V_s/r. $f(r,t)$ may be expressed as

$$f(r,t)=\frac{1}{r}\times \mathrm{Rect}_{r/V_s}(t)\times\left(\frac{V_s}{r}\right)^2\times t$$
$$=\frac{V_s^2}{r^3}\times \mathrm{Rect}_{r/V_s}(t)\times t, \tag{10}$$

where $\mathrm{Rect}_{\Delta}(t)$ is a rectangular window of unity height and duration Δ. The coupling term behavior is investigated in the Fourier domain. The Fourier transform of $f(r,t)$ is $F(r,\omega)$ where ω is the angular frequency

$$F(r,\omega)=\int_{-\infty}^{+\infty} f(r,t)e^{+i\omega t}\,dt=\frac{V_s^2}{r^3}\int_0^{r/V_s} te^{+i\omega t}\,dt. \tag{11}$$

$F(r,\omega)$ is estimated using an integration by parts that gives

$$F(r,\omega)=\frac{V_s}{\omega r^2}\left[\frac{V_s}{\omega r}e^{+i\omega r/V_s}-ie^{+i\omega r/V_s}-\frac{V_s}{\omega r}\right]. \tag{12}$$

It comes

$$F(r,\omega)=\frac{V_s^2}{\omega^2 r^3}\left[1-i\frac{\omega r}{V_s}-e^{-i\omega r/V_s}\right]e^{+i\omega r/V_s}. \tag{13}$$

For a given angular frequency ω, the behavior of $F(r,\omega)$ only depends on the distance r. Two particular cases must be treated independently:

(i) $r\to 0$,
(ii) $r\to\infty$.

In case (i), Eq. (13) simplifies by means of the third-order Taylor development of e^x for x close to zero:

$$e^x\simeq 1+x+\tfrac{1}{2}x^2+\tfrac{1}{6}x^3. \tag{14}$$

$F(r,\omega)$ may be written as

$$F_{r\to 0}(r,\omega)\simeq\frac{V_s^2}{\omega^2 r^3}\left[\frac{1}{2}\left(\frac{\omega r}{V_s}\right)^2-i\frac{1}{6}\left(\frac{\omega r}{V_s}\right)^3\right]e^{i\omega r/V_s}$$
$$=\frac{1}{2r}\left(1-i\frac{\omega r}{3V_s}\right)e^{i\omega r/V_s}. \tag{15}$$

The asymptotic expressions of its amplitude $A(r,\omega)$ and phase $\varphi(r,\omega)$ are

$$\varphi_{r\to 0}(r,\omega)\simeq\arctan\left(-\frac{\omega r}{3V_s}\right)+\frac{\omega r}{V_s}\simeq-\frac{\omega r}{3V_s}+\frac{\omega r}{V_s}$$
$$=\frac{2\omega r}{3V_s},$$
$$A_{r\to 0}(r,\omega)=\frac{1}{2r}. \tag{16}$$

The phase and amplitude asymptotic expressions indicate that the near field term propagates with a phase velocity equal to $3V_s/2$ and an r^{-1} decay. In case (ii), Eq. (13) may be written as

$$F_{r\to\infty}(r,\omega)\simeq\frac{V_s}{\omega r^2}e^{i(\omega r/V_s-\pi/2)}, \tag{17}$$

with

$$\varphi_{r\to\infty}(r,\omega)\simeq\frac{\omega r}{V_s}-\frac{\pi}{2},$$
$$A_{r\to\infty}(r,\omega)=\frac{V_s}{\omega r^2}. \tag{18}$$

In this case, the near field term propagates with the same phase velocity as the shear wave and a classical near field r^{-2} decay.

C. Discussion

The limit between the near field and the very near field occurs at a distance that may be defined as the distance r_0 for which the two asymptotic expressions of the phase delay are equal,

$$\varphi_{r\to\infty}(r_0,\omega)=\varphi_{r\to 0}(r_0,\omega), \tag{19}$$

$$\frac{\omega r_0}{V_s}-\frac{\pi}{2}=\frac{2\omega r_0}{3V_s}, \tag{20}$$

$$r_0=\frac{3V_s}{4f}=\frac{3\lambda_s}{4}, \tag{21}$$

where $2\pi f=\omega$ and λ_s is the shear wavelength. This characteristic distance is of great significance since it gives an indication of how the phase delay and amplitude laws behave in the near field. However, the directivity coefficient must be taken into account since it governs the spatial distribution of the coupling field.

For distances shorter than r_0, the amplitude behaves as r^{-1} and the phase velocity is $3V_s/2$. When the distance becomes larger than the wavelength the amplitude behaves as r^{-2} and the coupling term results in a wave propagating at

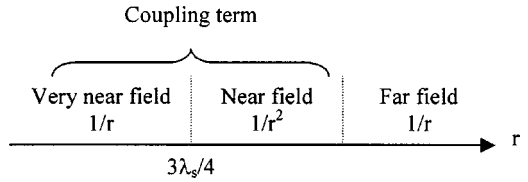


FIG. 2. The near field may be divided into a classical near field and a very near field in which a classical r^{-2} behavior and a r^{-1} behavior are observed respectively.

the shear velocity. This last point was already noticed by Aki and Richards. A simple interpretation is the following. When the wavelength is large compared to the distance r, then during the convolution of the emitted signal by the weighted ramp, all of the surface of the ramp contributes to the response. As we noticed in Sec. II A, the surface of the ramp is constant and the r^{-1} weighted factor forces an r^{-1} behavior. On the contrary, when the wavelength is short compared to the duration of the ramp, then only the height of the ramp significantly contributes to the amplitude of the response. Since the height of the ramp behaves as r^{-1}, its combination with the r^{-1} weighted factor leads to a global r^{-2} decrease.

These remarks are important because they highlight the complexity of the near field term behavior. Close to the source this term features the same characteristic r^{-1} decay as far field terms. This part of the field at distances shorter than $3\lambda_s/4$ might be called very near field (Fig. 2). In the classical near field at distances larger than the wavelength the decay reaches a more classical r^{-2} behavior.

The coupling term is particularly interesting in soft media ($\mu \ll \lambda$). Indeed, in this case the compressional term that is proportional to $1/\lambda$ can be neglected by comparison with the coupling and shear terms that are proportional to $1/\mu$. Furthermore, in the case of a point source, due to symmetry consider ations, the displacements are purely longitudinal on the axis parallel to the excitation. Thus the shear term contribution is zero on this axis. With a neglectable compressional term and a shear term equal to zero for symmetry considerations, the coupling term of the Green's function is predominant. Thus, surprisingly, the coupling term yields to a wave that propagates at a phase velocity equal to the shear velocity in the near field and one and a half times faster in the very near field. Moreover, though it propagates at the same velocity as the shear wave, this coupling wave is purely longitudinal on the axis of a point source.

III. COMPUTATION OF THE IMPULSE RESPONSE FROM A PISTON

In the previous section, we analyzed the Green's function corresponding to a free infinite solid, thus the displacement vector associated to the elastic waves that propagate in the solid due to a pointlike source located at the origin of the spatial coordinate system. In this section, we now consider an extended source located in the plane $z=0$ that works in the piston mode and generates a uniform source force $\mathbf{a}(0,0,a)$ over its surface S. It results from the previous expressions that the i component of the total displacement at any location $\mathbf{r}(x_1,x_2,x_3=z)$ in front of the vibrating piston can be written as

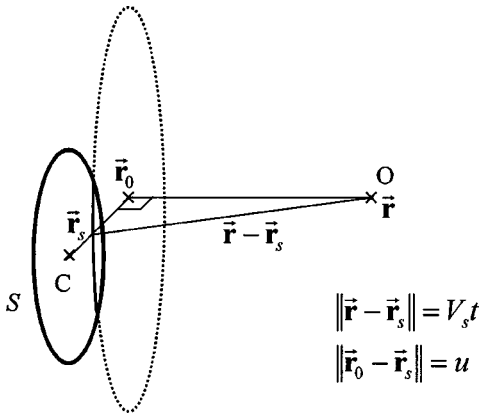


FIG. 3. Illustration of the new coordinate system after change of integration variable over the active surface of the disk.

$$u_{i3}(\mathbf{r},t) = a\int_S [G^c_{i3}(\mathbf{r}-\mathbf{r}_s,t) + G^s_{i3}(\mathbf{r}-\mathbf{r}_s,t) + G^{cs}_{i3}(\mathbf{r}-\mathbf{r}_s,t)]d^2\mathbf{r}_s = u^c_{i3}(\mathbf{r},t) + u^s_{i3}(\mathbf{r},t) + u^{cs}_{i3}(\mathbf{r},t). \quad (22)$$

We first consider the purely longitudinal and purely shear contributions to the total displacement. Looking at the Green's solution given by Eq. (2), we can observe that $G^c_{i3}(\mathbf{r},t)$ and $G^s_{i3}(\mathbf{r},t)$ have a very similar closed form expression. Using the notation $V_\alpha = V_{c|s}$ where α is either c (compressional waves) or s (shear waves), we can use a generic common formulation as follows:

$$G^{c|s}_{i3}(\mathbf{r},t) = \frac{1}{4\pi\rho V_\alpha^2}\frac{d^\alpha_i(\mathbf{r})}{r}\delta\left(t-\frac{r}{V_\alpha}\right),$$

where the directivity patterns $d^\alpha_i(\mathbf{r})$ are respectively given by $d^c_i(\mathbf{r}) = \gamma_i\gamma_3$ and $d^s_i(\mathbf{r}) = \delta_{i3} - \gamma_i\gamma_3$. It results therefore that the corresponding contributions to the total displacement vector are given by

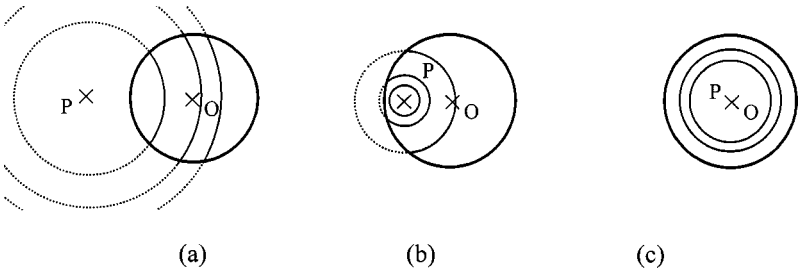
$$u^{c|s}_{i3}(\mathbf{r},t) = \frac{a}{4\pi\rho V_\alpha^2}\int_S \frac{d^\alpha_i(\mathbf{r}-\mathbf{r}_s)}{|\mathbf{r}-\mathbf{r}_s|}\delta\left(t-\frac{|\mathbf{r}-\mathbf{r}_s|}{V_\alpha}\right)d^2\mathbf{r}_s. \quad (23)$$

The computation of the integration over the surface S of the piston is based on two changes of integration variables.

The first change of integration variables consists in considering a description of the integration surface S using a cylindrical coordinate system (u,φ) centered at a point $\mathbf{r}_0$ that results from the projection of the observation point $\mathbf{r}$ in the plane $z=0$. In this new coordinate system, u is the distance between any point $\mathbf{r}_s$ and the new origin $\mathbf{r}_0$, and φ corresponds to the angle in the plane $z=0$, up to any arbitrary angle reference. Using this new coordinate system, we have $d^2\mathbf{r}_s \equiv u\,du\,d\varphi$ and $|\mathbf{r}-\mathbf{r}_s|^2 = u^2+z^2$. This first change of integration variable is illustrated in Fig. 3.

The second change of integration variables consists in replacing u by $\tau = |\mathbf{r}-\mathbf{r}_s|/V_\alpha$. As an immediate consequence, we have $u\,du = V_\alpha^2\tau\,d\tau$.

The last point is now to determine the new integration limits using the integration variables (τ,φ). By geometrical



(a) (b) (c)

FIG. 4. Integration over the vibrating surface (a) for an observation point outside the geometrical shadow of the disk (the arc grows and then decreases), (b) inside the geometrical shadow (the arc is a complete circle and then decreases), and (c) on the symmetry axis of the disk (the arc is a complete circle and then falls to an empty intersection with the vibrating disk).

interpretation, we can easily verify that the integration range over τ is $[t_{\min},t_{\max}]$, where $t_{\min}$ (resp. $t_{\max}$) is the minimum (resp. maximum) propagation time from $\mathbf{r}$ to any point on the surface of the piston S, calculated for the velocity V_α. In addition to that, we can observe that the directivity function $d_i^\alpha(\mathbf{r}-\mathbf{r}_s)$ only depends on the orientation of $\mathbf{r}-\mathbf{r}_s$; consequently this directivity only depends on φ. For simplicity, we do not change the notation and the directivity now becomes $d_i^\alpha(\varphi)$.

The transformations of the integral and the resulting new integration range are illustrated on Fig. 4.

After a few elementary simplifications, we obtain from Eq. (23) the following expression

$$u_{i3}^{c|s}(\mathbf{r},t)=\frac{a}{4\pi\rho V_\alpha}\int_\varphi\int_{t_{\min}}^{t_{\max}} d_i^\alpha(\varphi)\,\delta(t-\tau)\,d\tau\,d\varphi. \quad (24)$$

In this equation, the angular integration over φ is geometrically limited to the arc that results from the intersection between (i) the aperture surface of the piston S and (ii) a sphere of radius $V_\alpha t$ centered at the observation point $\mathbf{r}$.

Due to the Dirac distribution, the integration over τ is straightforward and we obtain

$$u_{i3}^{c|s}(\mathbf{r},t)=\begin{cases} 0, & \text{if } t<t_{\min} \text{ or } t>t_{\max},\\ \dfrac{a}{4\pi\rho V_\alpha}\displaystyle\int_\varphi d_i^\alpha(\varphi)\,d\varphi, & \text{otherwise.}\end{cases} \quad (25)$$

This quite simple equation allows vs to interpret the purely longitudinal or purely transverse contributions to the total displacement vector as an angular averaging of the directivity function $d_i^\alpha(\varphi)$.

We now consider the particular case of a circular piston of radius R located in the plane $z=0$ and centered at the origin of our spatial coordinate system. As assumed above, this piston generates a uniform source force $\mathbf{a}(0,0,a)$. In these conditions, we show here that we can derive a closed form expression in any situation.

Due to the circular symmetry of the problem, we can consider without loss of generality that the observation point $\mathbf{r}$ is located at $x_1=0$, $x_2=y\geqslant 0$ and $x_3=z>0$. The angular integration variable φ is the angle of any point on the vibrating surface S with respect to the second axis of our spatial coordinate system. We now distinguish between the different components of the total displacement field to calculate. For the following discussion, we first introduce the following temporal parameters:

$$t_0=z/V_\alpha, \quad t_1=\sqrt{(R-y)^2+z^2}/V_\alpha,$$

and

$$t_2=\sqrt{(R+y)^2+z^2}/V_\alpha. \quad (26)$$

Concerning the first component of the displacement, we easily observe that the two directivity functions $d_1^\alpha(\varphi)$ are both proportional to γ_1, thus to the first component of $\mathbf{r}-\mathbf{r}_s$. In other words, these directivity functions vary as $\sin\varphi$. We also observe that the angular integration given by Eq. (25) always corresponds to a symmetrical domain of φ. We can immediately deduce that the first component of the total displacement is always 0. This result is not surprising and is a direct consequence of the symmetry of the problem.

Similarly to the previous case, the two directivity functions $d_2^\alpha(\varphi)$ are both proportional to γ_2, thus to $\cos\varphi$. We have different regimes, depending on the position of the observation point with respect to the aperture of the vibrating surface. The three characteristic situations are as follows:

(i) The observation point is located on the symmetry axis of the radiating disk, such that we have $y=0$; in this case, $t_{\min}$ and $t_{\max}$ are respectively equal to t_0 and t_2. For t less than t_0 or larger than t_2, the intersection of the sphere of radius $V_\alpha t$ and the vibrating disk is empty and the resulting displacement is 0. Inside the temporal interval $[t_0,t_2]$, we immediately see from the geometrical interpretation that this intersection is a complete circle, whose radius increases with t. Anyway, the angular integration must be calculated from $-\pi$ to $+\pi$ and the resulting displacement is also 0.

(ii) The observation point is located inside the geometrical shadow of the vibrating disk, such that $0<y<R$; as in the previous case, $t_{\min}$ and $t_{\max}$ are respectively equal to t_0 and t_2, but the value of t_1 is also involved in the expression of the displacement vector. As above, the total displacement is 0 outside the temporal interval $[t_0,t_2]$. Inside this interval and for t less than t_1, the intersection we have to consider is a complete circle as in (i), and the resulting displacement is 0 again. A different regime arises for t larger than t_1 since the intersection is now a portion of circle, instead of the complete circle as above. Consequently, the integration over φ no more vanishes. More precisely, we have $\gamma_2=\cos\varphi\sqrt{1-z^2/V_\alpha^2t^2}$ and $\gamma_3=z/V_\alpha t$, and the directivity functions are given by

$$d_2^\alpha(\varphi)=\pm\cos\varphi\,\frac{z}{V_\alpha t}\sqrt{1-\frac{z^2}{V_\alpha^2t^2}},$$

where the $+$ (resp. $-$) sign corresponds to the purely longitudinal (resp. transverse) component. Finally, the angular integration reduces to a symmetrical interval $[-\varphi_m,+\varphi_m]$, where φ_m is given by the following relationship:

$$\varphi_m = \cos^{-1}\left(\frac{V_\alpha^2 t^2 - z^2 + y^2 - R^2}{2y\sqrt{V_\alpha^2 t^2 - z^2}}\right).$$

From these equations, we can give the closed form expression for the total displacement in this particular case as

$$u_{23}^{c|s}(\mathbf{r},t) = \pm\frac{a}{2\pi\rho V_\alpha}\frac{z}{V_\alpha t}\sqrt{1-\frac{z^2}{V_\alpha^2 t^2}}\sin\varphi_m. \tag{27}$$

(*iii*) The observation point is located outside the geometrical shadow of the vibrating disk, such that $y>R$; in this case, $t_{\min}$ and $t_{\max}$ are respectively equal to t_1 and t_2. Inside the temporal interval $[t_1,t_2]$, the intersection is now a portion of circle, and never simplifies to the complete circle. As in the previous case, the angular integration over φ reduces to $[-\varphi_m,+\varphi_m]$, where the previous expression given for φ_m remains valid. Consequently, the expression given by Eq. (27) is still usable.

The last case of interest is the third component of the displacement field. In this last case, the directivity functions $d_3^\alpha(\varphi)$ only depend on γ_3 that does not vary with φ. As an immediate consequence, the angular integration over φ reduces to the computation of the length of the arc that describes the intersection of the sphere of radius $V_\alpha t$ with the radiating disk. Up to this directivity coefficient, the closed form expression of the third component of the displacement field reduces to same form as the standard impulse response of a disk in a fluid.

The previous results can be summarized as follows.

(i) The first component of the displacement field is 0, for any observation point $\mathbf{r}$ and any observation time t

(ii) The second component of the displacement field is

$$u_{23}^{c|s}(\mathbf{r},t) = \begin{cases} 0, & \text{if } t<t_1, \\ \pm\dfrac{a}{2\pi\rho V_\alpha}\dfrac{z}{V_\alpha t}\sqrt{1-\dfrac{z^2}{V_\alpha^2 t^2}}\sin\varphi_m, & \text{if } t_1<t<t_2, \\ 0, & \text{if } t>t_2, \end{cases} \tag{28}$$

where the + (resp. −) sign corresponds to the purely longitudinal (resp. transverse) component.

(iii) The third component of the displacement field is for $y<R$

$$u_{33}^{c|s}(\mathbf{r},t) = \begin{cases} 0, & \text{if } t<t_0, \\ \dfrac{a}{2\rho V_\alpha}\left([1-]\dfrac{z^2}{V_\alpha^2 t^2}\right), & \text{if } t_0<t<t_1, \\ \dfrac{a}{2\pi\rho V_\alpha}\left([1-]\dfrac{z^2}{V_\alpha^2 t^2}\right)\varphi_m, & \text{if } t_1<t<t_2, \\ 0, & \text{if } t>t_2. \end{cases} \tag{29}$$

In the other case $y>R$, the final expression looks similar and is given by

$$u_{33}^{c|s}(\mathbf{r},t) = \begin{cases} 0, & \text{if } t<t_1, \\ \dfrac{a}{2\pi\rho V_\alpha}\left([1-]\dfrac{z^2}{V_\alpha^2 t^2}\right)\varphi_m, & \text{if } t_1<t<t_2,. \\ 0, & \text{if } t>t_2 \end{cases} \tag{30}$$

In these expressions, the quantity in brackets is present only in the case of the purely transverse component of the displacement.

In the previous computation steps, we finally observe that the approach is exactly the same as the one used by multiple authors in the computation of the diffraction of a disk piston immersed in water. Our method uses the same transformations; the only difference is that we now have a angular-dependent term to integrate over the surface of the radiating disk, which describes the directivity of the elastic field generated by a pointlike source in a free infinite solid. Anyway, this directivity term can be evaluated in a closed form without any particular difficulties, and we finally obtain a generalization of the impulse formalism previously developed for radiating disks in a fluid.[19–21] We are now interested in the computation of the contribution of the coupling term. From Eq. (22), this contribution is given by

$$u_{i3}^{cs}(\mathbf{r},t) = a\int G_{i3}^{cs}(\mathbf{r}-\mathbf{r}_s,t)\,d^2\mathbf{r}_s,$$

where $G_{i3}^{cs}(\mathbf{r},t)$ is given by Eq. (2). By change of integration variables, the surface of the vibrating piston is described by a cylindrical coordinate system (u,φ), centered at the projection of the observation point $\mathbf{r}$ on the plane $z=0$. The angle φ is considered with respect to the second axis of our initial coordinate system.

Using this new representation of the vibrating surface, we have

$$d^2\mathbf{r}_s = u\,du\,d\varphi, \quad |\mathbf{r}-\mathbf{r}_s| = \sqrt{z^2+u^2},$$

$$\gamma_1 = \frac{u\sin\varphi}{\sqrt{z^2+u^2}}, \quad \gamma_2 = \frac{u\cos\varphi}{\sqrt{z^2+u^2}}, \quad \text{and} \quad \gamma_3 = \frac{z}{\sqrt{z^2+u^2}}.$$

Using these new variables, the integration range for u is simply $[0,y+R]$ if y is less than R, and $[y-R,y+R]$ otherwise. The angular integration range for φ is a little more complex. If u is less than $R-y$, the angular integration has to be calculated over a complete circle, thus $[-\pi,+\pi]$. Otherwise, only a portion of the circle must be considered and the corresponding integration is limited to the symmetrical interval $[-\varphi_m(u),+\varphi_m(u)]$, where $\varphi_m(u)$ is defined by

$$\varphi_m(u) = \cos^{-1}\left(\frac{u^2+y^2-R^2}{2uy}\right).$$

Clearly, these two regimes only exist in the particular case $y<R$, corresponding to an observation point that is located in the geometrical shadow of the vibrating disk. The contribution of the coupling ramp to the total displacement field can finally be written in the following form:

$$u_{i3}^{cs}(\mathbf{r},t)=\frac{a}{4\pi\rho}\int_u\int_\varphi\frac{3\gamma_i\gamma_3-\delta_{i3}}{(z^2+u^2)^{3/2}}u\,du\,d\varphi \times\int_{\sqrt{z^2+u^2}/V_c}^{\sqrt{z^2+u^2}/V_s}\tau\delta(t-\tau)d\tau. \quad (31)$$

The integral over τ is a function of u only and is given by $th(\sqrt{z^2+u^2})$, where the auxiliary function $h(v)$ is defined by

$$h(v)=H(v/V_s-t)-H(v/V_c-t).$$

In the most general case, the expression given by Eq. (31) is too complex and we cannot obtain a closed form solution. Anyway, for an observation point located on the symmetry axis of the radiating disk ($y=0$), the expression given above can be simplified since the angular integration always corresponds to a complete circle, thus $[-\pi,+\pi]$.

In this particular case, the first and second components of the displacement vector are respectively proportional to the average of $\cos\varphi$ and $\sin\varphi$ between $-\pi$ and $+\pi$; these two components clearly vanish and reduce to 0. For the third component, we easily observe that the expression to integrate over the radiating surface does not depend on φ and the corresponding integration is straightforward. The integration over u covers the range $[0,R]$ and we can easily verify that the formulation only depends on $v=\sqrt{z^2+u^2}$. Thus, the expression given by Eq. (31) reduces to

$$u_{33}^{cs}(\mathbf{r},t)=\frac{at}{2\rho}\int_0^R\frac{h(v)}{v^3}\left(\frac{3z^2}{v^2}-1\right)u\,du =\frac{at}{2\rho}\int_z^{\sqrt{z^2+R^2}}h(v)\left(\frac{3z^2}{v^4}-\frac{1}{v^2}\right)dv.$$

This last expression can be simplified based on the observation that the contribution of $h(v)$ is either 0 or 1, depending on the value of v compared to V_st and V_ct. More precisely, $h(v)$ is nonzero only if v is larger than V_st and less than V_ct. As an immediate consequence, the last expression given above can be calculated in a closed form as follows:

$$u_{33}^{cs}(\mathbf{r},t)=\frac{at}{2\rho}\int_{m_1}^{m_2}\left(\frac{3z^2}{v^4}-\frac{1}{v^2}\right)dv =\frac{at}{2\rho}\left[\frac{1}{m_2}-\frac{1}{m_1}\right]-\frac{atz^2}{2\rho}\left[\frac{1}{m_2^3}-\frac{1}{m_1^3}\right], \quad (32)$$

where the interval $[m_1,m_2]$ is defined by $[z,\sqrt{z^2+R^2}]\cap[V_st,V_ct]$.

To calculate the final expression of the total displacement vector on the symmetry axis of the radiating disk, it is necessary to distinguish between the two following situations:

$$\text{Case I:}\quad \frac{z}{V_c}<\frac{\sqrt{z^2+R^2}}{V_c}<\frac{z}{V_s}<\frac{\sqrt{z^2+R^2}}{V_s},$$

$$\text{Case II:}\quad \frac{z}{V_c}<\frac{z}{V_s}<\frac{\sqrt{z^2+R^2}}{V_c}<\frac{\sqrt{z^2+R^2}}{V_s}.$$

The basic reason for this distinction is that we have to discuss the different contributions depending on the value of the observation time t with respect to some limits that depend on z, $\sqrt{z^2+R^2}$, V_c, and V_s. A common result is that only the third component of the displacement vector is nonzero, due to the symmetry of the radiating disk.

After a few computation steps that we do not give in detail, we finally obtain from Eqs. (29) and (32) the same expression for the third component of the displacement vector $u_{33}(\mathbf{r},t)$ in the two cases mentioned above:

$$u_{33}(\mathbf{r},t)=\begin{cases}0, & \text{if } t<z/V_c,\\ \dfrac{a}{2\rho V_c}, & \text{if } z/V_c<t<\sqrt{z^2+R^2}/V_c,\\ \dfrac{aR^2t}{2\rho(z^2+R^2)^{3/2}}, & \text{if } \sqrt{z^2+R^2}/V_c<t<\sqrt{z^2+R^2}/V_s,\\ 0, & \text{if } t>\sqrt{z^2+R^2}/V_s.\end{cases} \quad (33)$$

IV. SIMPLIFIED GREEN'S FUNCTION AND IMPULSE RESPONSE IN A SEMI-INFINITE MEDIUM

Based on the results presented in the preceding sections, a simplified Green's function for a semi-infinite homogeneous and isotropic medium is proposed. The longitudinal components of the on-axis exact and simplified Green's functions are compared. These comparisons are obtained for soft media. Finally, the on-axis impulse response associated with a circular radiator is discussed.

A. Simplified Green's function expression

The simplified Green's function in a semi-infinite medium differs from the classical Green's function in an infinite medium only by a multiplicative factor of 2. The origin of this factor is related to the presence of the plane interface which acts as a mirror and yields a secondary source point image of the original source by symmetry through the plane interface. In our case, the original source is located on the interface, such that the secondary source coincides with the original one. It is expressed as

$$G_{ij}(\mathbf{r},t)=G_{ij}^c(\mathbf{r},t)+G_{ij}^s(\mathbf{r},t)+G_{ij}^{cs}(\mathbf{r},t), \quad (34)$$

where

$$G_{ij}^c(\mathbf{r},t)=\frac{1}{2\pi\rho V_c^2}\gamma_i\gamma_j\frac{1}{r}\delta\left(t-\frac{r}{V_c}\right),$$

$$G_{ij}^s(\mathbf{r},t)=-\frac{1}{2\pi\rho V_s^2}(\gamma_i\gamma_j-\delta_{ij})\frac{1}{r}\delta\left(t-\frac{r}{V_s}\right), \quad (35)$$

$$G_{ij}^{cs}(\mathbf{r},t)=\frac{1}{2\pi\rho}(3\gamma_i\gamma_j-\delta_{ij})\frac{1}{r^3}\int_{r/V_c}^{r/V_s}\tau\delta(t-\tau)d\tau.$$

Assumptions that justify the use of the simplified Green's function have been given by Lhemery.[22] The main assumption is that the mode conversions at the free surface can be

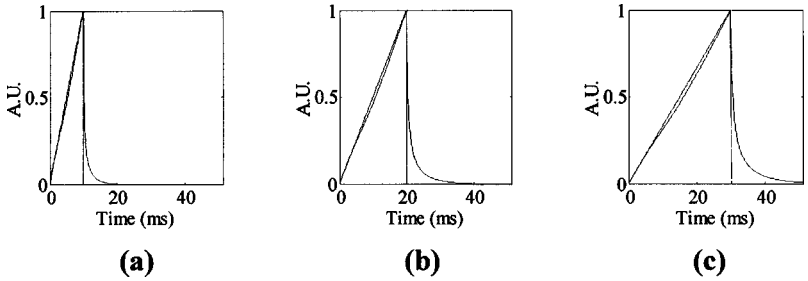


FIG. 5. Comparison between the axial exact (dotted line) and simplified (plain line) Green's functions for a point situated on the axis of the point load source at depths of (a) 50 mm, (b) 100 mm, and (c) 150 mm.

neglected. However, Lhemery neglected the coupling term in his model. In this section, the simplified Green's function expression including the coupling term is used in order to take into account the contribution of the temporal ramp which is particularly important in soft media.

B. On-axis Green's functions

The longitudinal exact Green's function on the symmetry axis of a point load source as given by Gakenheimer is compared to the simplified Green's function. A point load source refers to an excitation perpendicular to the free surface. The expressions used for the calculation of the exact Green's functions are given in the Appendix. Calculations are made at depths of 50, 100, and 150 mm in a soft medium where the compressional and shear velocities are respectively 1500 and 5 m/s. Results are shown in Fig. 5. The two Green's functions feature a ramp between the arrival times of the compressional and shear waves. Since the compressional velocity is large compared to the shear velocity, the ramps start at a time close to zero and end respectively at times 10, 20, and 30 ms. The main difference between the exact and simplified Green's functions arises after the end of the ramp and results in a tail that is not present in the simplified Green's function. This tail is attributed to phenomena that are not taken into account in the simplified model such as mode conversions at the interface. The conversion of surface waves into shear waves seems the more likely to occur. These differences show the limits of the simplified model.

The variations of the phase delay and amplitude of the exact and simplified Green's functions versus depth (1–100 mm) have been investigated at a frequency of 150 Hz. Given the shear velocity of 5 m/s, the shear wavelength λ_s is 33 mm. Comparisons are shown in Fig. 6. The separation between the near field and the very near field should occur around the characteristic depth $3\lambda_s/4$ (see Sec. II C) of 25 mm marked by black arrows. The amplitude is plotted in a log-log scale. The phase delay laws [Fig. 5(a)] are in very good agreement and verify the theoretical predictions: the phase velocity varies from 7.5 m/s at shallow depth and asymptotically reaches 5 m/s for large depths. The amplitude [Fig. 5(b)] is perturbed close to source but keeps the same behavior whatever the Green's function with a r^{-1} decay for depth shorter than 25 mm and a r^{-2} decay for larger depths.

In a first approximation, the simplified model and the exact solution are in good agreement on the axis of the source. Thus assumptions made in the present paper seem to be valid. Mode conversions at the interface can be neglected.

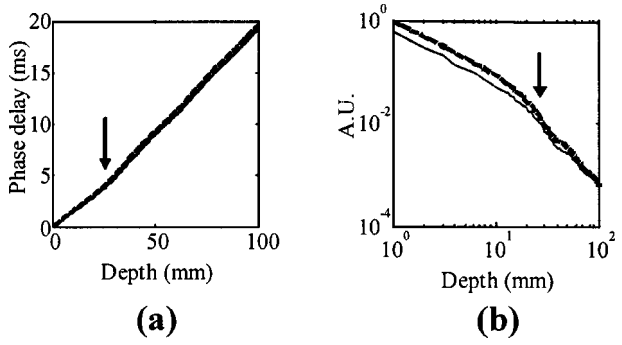


FIG. 6. (a) Phase delay and (b) amplitude behavior of the exact (dotted line) and simplified (plain line) Green's functions at 150 Hz on the axis of the source; the very near field extends till 25 mm (black arrows).

C. On-axis impulse response

From now on, only point sources will be studied. In this paragraph, the case of circular vibrators is introduced. The impulse response is helpful to analyze the diffraction effects that superimpose to the near field coupling effects described in the preceding section. For a circular radiator, Eq. (33) gives an analytical expression of the impulse response. In the case of soft semi-infinite media, the equation simplifies to

$$u_{33}(z,t)=\begin{cases}\dfrac{aR^2t}{\rho(z^2+R^2)^{3/2}}, & \text{if } 0\leqslant t<\dfrac{\sqrt{z^2+R^2}}{V_s},\\ 0, & \text{if } \dfrac{\sqrt{z^2+R^2}}{V_s}\leqslant t,\end{cases} \tag{36}$$

where u_{33} is the displacement observed in the z direction for a load of amplitude a applied in the z direction.

The impulse response analytical expression takes into account both diffraction and coupling effects for circular sources on homogeneous and isotropic half-space media. The noticeable point is that the shape of the impulse response is a temporal ramp similar to the coupling term of the Green's function in an infinite medium.

V. EXPERIMENTAL INVESTIGATION

Experimental result impulse responses obtained using transient elastography are compared with theoretical impulse responses calculated with the simplified Green's function model. Transient elastography has been developed in order to measure tissue stiffness *in vivo*. The technique is part of other elastographic techniques that essentially differ by the kind of perturbation which is applied to the medium and by the imaging system used to estimate displacements within the medium. Transient elastography can be used to follow low frequency transient waves in tissues.

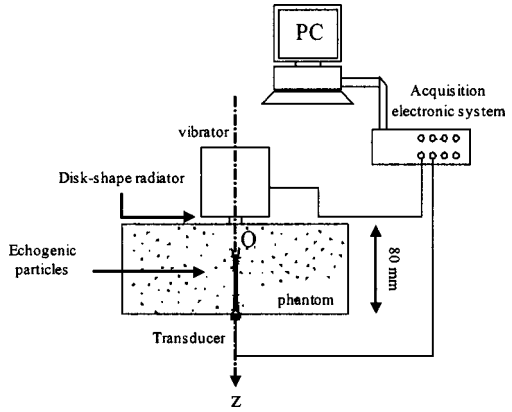


FIG. 7. Experimental setup.

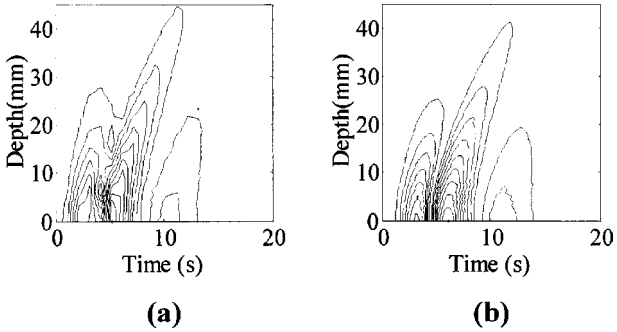


FIG. 8. Experimental (a) and theoretical (b) displacement fields for a surface source diameter of 20 mm.

A. Transient elastography

The purpose of transient elastography is to measure the elasticity of soft tissues. It involves both a low-frequency system and a high-frequency (ultrasonic) system. A single ultrasonic transducer or an array of transducers is used to measure the displacements parallel to the ultrasonic beam. The displacements are induced in the soft medium by a low-frequency vibrator that generates shear waves whose velocity V_s is directly related to the Young's modulus E. In a purely elastic soft medium where dissipation can be ignored and the Poisson ratio is close to $\frac{1}{2}$,

$$E = 3\rho V_s^2, \tag{37}$$

where ρ is the mass density. Since the displacements are measured versus time and depth, this technique can be used to verify the theoretical expression of the impulse response in a semi-infinite medium presented in Sec. IV.

B. Experimental setup

The experimental setup is described in Fig. 7. The medium is a tissue-mimicking gel-based phantom. The phantom is homogeneous and its size is $80\times100\times200$ mm. It is made with gelatine (4%) and agar (3%) powders mixed in hot water (80 °C). Five percent propanol-1 is added to the solution as a conservative agent. Agar powder is used as echogenic particles while gelatine concentration is related to elasticity. It should be noted that the gelatine and agar concentrations were chosen in order to obtain a phantom in which the shear velocity is 5 m/s.

Seven circular radiators of various diameters (ϕ: 1, 2, 5, 10, 12, 15 and 20 mm) can be fixed to the low-frequency vibrator (Bruël & Kjær minishaker type 4810) and placed on the top interface of the phantom. A single 7-mm-diam transducer acts as a pulse-echo emitter receiver. Its center frequency is 5.0 MHz and it is focused at 35 mm. The transducer and circular radiator are placed on the same axis on each side of the phantom. The free surface on which the circular radiator is placed is at $z=0$ mm.

A dedicated single channel electronic system controls both the ultrasonic pulsed echo mode and the low-frequency excitation. Ultrasonic signals are sampled at 50 MHz. They are recorded using a 9-bit digitizer with 2 Mbytes random access memory (RAM). Ultrasonic signals (A-lines) are acquired at a repetition frequency of 1000 Hz and transferred to the computer. An acquisition sequence consists in sending the low-frequency excitation to the minishaker while the electronic system operates in pulsed echo mode and stores A-lines at a repetition frequency of 1000 Hz.

The displacements are measured using the standard cross-correlation technique[23] between $z=0$ and $z=50$ mm with a 1-μm precision. Only longitudinal displacements (along axis z) induced on the axis of the source are measured.

The low-frequency excitation signal is a period of sinusoid at 150 Hz. The peak to peak amplitude of the vibration is 1 mm. In transient elastography, the excitation frequency typically varies between 50 and 200 Hz. The choice of a quite high frequency is related to the finite size of the medium. Since we do not want to be sensitive to boundary conditions, a short wave length is necessary. Though high frequencies yield to high attenuation, a compromise was found at 150 Hz.

C. Experimental and theoretical results

Theoretical displacement fields are computed using the impulse response analytical expression of Eq. (36). The medium shear velocity and compressional velocity are respectively of 5 and 1500 m/s. Experimental and theoretical transient displacement fields obtained with the 20-mm-diam piston are plotted in Fig. 8 as a function of depth and time. The theoretical Green's functions have been convoluted by the displacement of the surface deduced from the experimental displacement field.

A comparison of the experimental and theoretical displacement fields is performed in the Fourier domain. The variations of the phase delay and amplitude of the on-axis longitudinal displacement versus depth are plotted in Fig. 9 for the seven different circular vibrators. The investigation depth is restricted between 0 and 40 mm in order to avoid boundary effects (such as reflection) at the bottom interface. We shall see that the missing 40 mm are of the order of one wavelength. For the smaller diameters (1, 2, and 5 mm), the experimental displacements above a certain depth were below the precision of the displacement estimator. Thus their measurement was not possible on the whole depth.

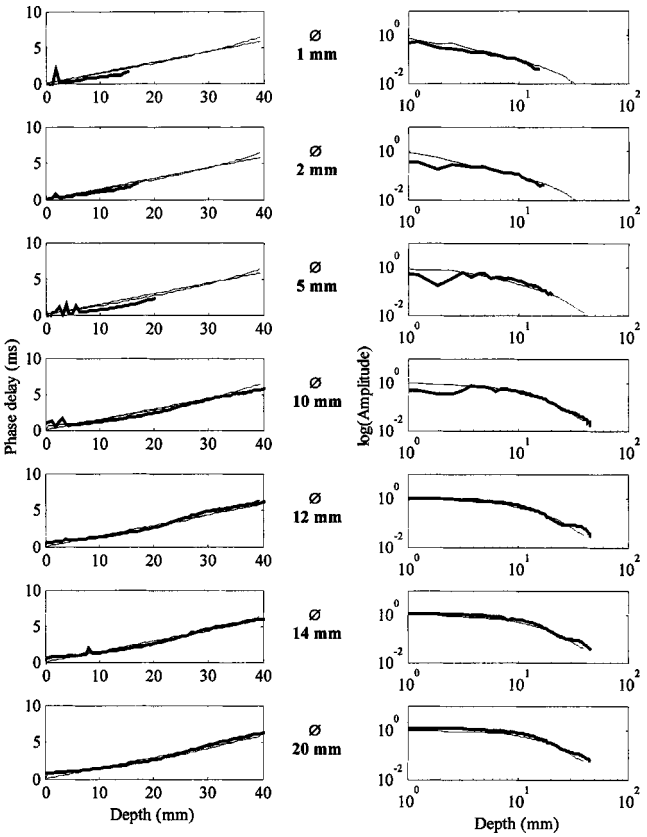


FIG. 9. Variations of the phase delay and amplitude of the on-axis longitudinal displacement versus depth for circular vibrators diameters ranging from 1 to 20 mm.

D. Discussion

Experimental and theoretical results are in very good agreement whatever the circular diameter used. The finite size of the experimental medium does not affect the results. The displacements close to the bottom interface are not taken into account because of the reflections that occur in this region of the medium. The noticeable point is that these results reveal both diffraction effects due to the extended pistonlike source and coupling effects. Coupling effects are more important for small pistons while diffraction effects dominate as the piston diameter increases.

Investigating coupling effects is complicated because as the diameter of the piston decreases, the amplitude of the displacements induced in the medium tends to go beyond the resolution of the system. Moreover, experimental measurements with small diameter disks are difficult to conduct because the alignment of the transducer and vibrator axes becomes a major issue and is hard to obtain. Furthermore, the ultrasonic beam lateral resolution $\lambda f/D$ is 1.2 mm, which is insufficient close to small diameter disks.

The amplitude observed with the 1-mm-diam vibrator for depth shallower than 10 mm decreases as $1/r$ which is consistent with the coupling term behavior associated with a point source as mentioned in Sec. II B, though results obtained with such a small vibrator are subject to biases. Diffraction effects result in the presence of a threshold in the amplitude variations observed at shallow depths in the case of large vibrators. Above the depth of ~20 mm, the amplitude decreases as r^{-2} which is characteristic of the coupling term behavior in the near field. The theoretical and experimental phase delays are in good agreement.

The simplified Green's function can be used to deduce the elasticity of soft media from the displacement field. The case of arbitrary shaped vibrators may be handled using numerical calculations developed in Sec. III. Diffraction and coupling effects in dynamic elastography can be efficiently compensated to avoid biases that can be due either to the finite dimension of the source or even to the immediate proximity of the source. In this last case the shear velocity may be overestimated by a factor $\frac{3}{2}$.

VI. CONCLUSION

In conclusion, the results presented in this paper clearly demonstrate that a longitudinally polarized wave propagating at the velocity of the shear wave can be observed in the near field of a point source. The amplitude and phase behaviors of this coupling wave vary as a function of the distance to the source. In the very near field region the amplitude behaves as r^{-1} and the velocity of the wave is one and a half times the shear velocity. For larger distances, the amplitude behaves more classically as r^{-2} and the wave propagates at the shear velocity. The case of circular vibrators was investigated using simplified Green's functions and impulse response calculations.

Comparisons have shown that theoretical results and experimental data obtained with transient elastography are in very good agreement. Thus, by taking in account the diffraction and coupling effects, transient elastography can be used to measure the shear velocity in soft media. The theoretical approach described in the paper can be used to compensate both diffraction and coupling effects observed in dynamic elastography. However, in this work the medium is assumed isotropic and homogeneous, which is quite restrictive and the study should be extended to the case of more complex media, such as anisotropic media or viscous media. Heterogeneous media such as layered media are also of great importance.

ACKNOWLEDGMENTS

The authors would like to thank Daniel Royer, Jean-Michel Hasquenoph, Isabelle Ménier, and the Ménier family for their helpful direct or indirect contributions to this work.

APPENDIX: GAKENHEIMER AND MIKLOWITZ EXACT GREEN'S FUNCTION

The exact Green's function along the point load source axis z was calculated by Gakenheimer and Miklowitz. It associates the displacements at a certain location on the z axis to the normal load applied at position O. It is expressed as the sum of a compression term G^c and a shear term G^s:

$$G_{zz} = G^s_{zz} + G^c_{zz}. \tag{A1}$$

The compression and shear terms depend on the shear modulus μ,

$$G_{zz}^{s}=-\frac{V_c}{\pi\mu z^2}\frac{\partial}{\partial t}\left[H(t-t_s)t\frac{k_2^s k_3^s}{k_5^s}\right],$$
$$G_{zz}^{c}=\frac{V_c}{2\pi\mu z^2}\frac{\partial}{\partial t}\left[H(t-t_c)t\frac{k_1^c k_3^c}{k_5^c}\right], \tag{A2}$$

in which μ is the shear modulus, V_c is the compression velocity, z is the depth, H is the Heavyside function and t_c and t_s are respectively the arrival times of the compression and shear waves. k symbols are defined by

$$k_1^s=\gamma^2\left(\frac{2t^t}{t_s^2}-1\right),\quad k_1^c=\left(\frac{2t^t}{t_c^2}-1\right)+\gamma^2,$$
$$k_2^s=\gamma^2\left(\frac{t^t}{t_s^2}-1\right),\quad k_2^c=\left(\frac{t^t}{t_c^2}-1\right),$$
$$k_3^s=\sqrt{\gamma^2\left(\frac{t^t}{t_s^2}-1\right)+1},\quad k_3^c=\frac{t}{t_c}, \tag{A3}$$
$$k_4^s=\gamma\frac{t}{t_s},\quad k_4^c=\sqrt{\frac{t^t}{t_c^2}-1+\gamma^2},$$
$$k_5^s=(k_1^s)^2-4k_2^s k_3^s k_4^s,\quad k_5^c=(k_1^c)^2-4k_2^c k_3^c k_4^c,$$

where $\gamma=V_c/V_s$. This solution is easily computable.

[1] R. M. Lerner, S. R. Huang, and K. J. Parker, "Sonoelasticity images derived from ultrasound signals in mechanically vibrated tissues," Ultrasound Med. Biol. **16**(3), 231–239 (1990).

[2] K. J. Parker, S. R. Huang, and R. M. Lerner, "Tissue response to mechanical vibrations for sonoelasticity imaging," Ultrasound Med. Biol. **16**(3), 241–246 (1990).

[3] F. Lee, J. P. Bronson, R. M. Lerner, K. J. Parker, S. R. Huang, and D. J. Roach, "Sonoelasticity Imaging: Results in Vitro Tissue Specimens," Radiology, 237–239 (1991).

[4] L. Gao, K. J. Parker, R. M. Lerner, and S. F. Levinson, "Imaging of the elastic properties of tissue—a review," Ultrasound Med. Biol. **22**(8), 959–977 (1996).

[5] R. Muthupillai, D. J. Lomas, P. J. Rossman, J. F. Greenleaf, A. Manduca, and R. L. Ehman, "Magnetic resonance elastography by direct visualization of propagating acoustic strain waves," Science **269**, 1854–1857 (1995).

[6] V. Dutt, R. R. Kinnick, and J. F. Greenleaf, "Acoustic Shear Wave Displacement Measurement using Ultrasound," IEEE Ultrasonics Symposium **2**, 1185–1193 (1996).

[7] S. Catheline, F. Wu, and M. Fink, "A solution to diffraction biases in sonoelasticity: The acoustic impulse technique," J. Acoust. Soc. Am. **105**, 2941–2950 (1999).

[8] S. Catheline, J. L. Thomas, F. Wu, and M. Fink, "Diffraction field of a low-frequency vibrator in soft tissues using transient elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**(4), 1013–1020 (1999).

[9] L. Sandrin, M. Tanter, J. L. Gennisson, S. Catheline, and M. Fink, "Shear elasticity probe for soft tissues with 1D transient elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**(4), 436–446 (2002).

[10] L. Sandrin, S. Catheline, M. Tanter, X. Hennequin, and M. Fink, "Time-Resolved Pulsed Elastography," Ultrason. Imaging **21**, 259–272 (1999).

[11] L. Sandrin, M. Tanter, S. Catheline, and M. Fink, "Shear Modulus Imaging with 2D Transient Elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**(4), 426–435 (2002).

[12] G. F. Miller and H. Pursey, "The field and radiation impedance of mechanical radiators on the free surface of a semi-infinite isotropic solid," Proc. R. Soc. London, Ser. A **233**, 521–541 (1954).

[13] K. Kawashima, "Theory and numerical calculation of the acoustical field produced in metal by an electromagnetic ultrasonic transducer," J. Acoust. Soc. Am. **60**, 1089–1099 (1975).

[14] H. Djelouah and J. C. Baboux, "Transient ultrasonic fields radiated by circular transducers in a solid medium," J. Acoust. Soc. Am. **92**, 2932–2941 (1992).

[15] J. C. Baboux and R. Kazys, "Analysis of the transient ultrasonic fields radiated in solids by circular and annular sources," J. Acoust. Soc. Am. **92**, 2942–2951 (1992).

[16] E. M. Timanin, "Displacement Field Produced by a Surface Source of Vibrations in a Layered Biological Tissue," Acoust. Phys. **48**(1), 98–104 (2002).

[17] B. N. Klochkov, "Near-field of a Low-Frequency Source Source of Forced Vibration on a Layered Biological Tissue," Acoust. Phys. **48**(1), 70–76 (2002).

[18] D. C. Gakenheimer and J. Miklowitz, "Transient excitation of an elastic half space by a point load travelling on the surface," J. Appl. Mech., 505–515 (1969).

[19] A. E. H. Love, *The Mathematical Theory of Elasticity* (Dover, New York, 1944), p. 305.

[20] J. D. Achenbach, *Wave Propagation in Elastic Solids* (North Holland, Amsterdam, 1973), p. 100.

[21] P. G. R. Aki, *Quantitative Seismology, Theory and Methods* (Freeman, San Francisco, 1980), Vol. 1, Chap. 4, p. 74.

[22] A. Lhémery, "A model for the transient ultrasonic fields radiated by an arbitrary loading in a solid," J. Acoust. Soc. Am. **96**, 3776–3786 (1994).

[23] J. A. Jensen, *Estimation of Blood Velocities Using Ultrasound* (Cambridge U.P., Cambridge, 1996), Chap. 8.

Acoustic cavity modes in lens-shaped structures

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

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

(Received 5 February 2003; accepted for publication 6 October 2003)

A method for solving exactly the Helmholtz equation in parabolic rotational coordinates is presented using separability of the eigenfunctions and the Frobenius power series expansion technique. Two examples of interest in acoustics are considered and analyzed quasianalytically: The acoustic pressure in a cavity defined by two paraboloids (forming a lens-shaped structure) with (I) rigid wall boundary conditions and (II) pressure-release boundaries. The rigid-wall (pressure-release) acoustic enclosure problem is a Neumann (Dirichlet) boundary condition problem. In both cases, eigenfunctions and eigenmodes are calculated and the shape dependence of the eigenvalue for the ground state is examined. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1630996]

PACS numbers: 43.20.Ks [MO] Pages: 84–90

I. INTRODUCTION

The Helmholtz equation is known to be separable in 11 coordinate systems including the parabolic rotational coordinate system (PRC).[1,2] In this paper, an exact method for determining the solutions of the Helmholtz equation in a closed region formed by two paraboloids is presented subject to either Dirichlet or Neumann boundary conditions.

The method described is applied to two problems of interest in acoustics: The pressure field eigenstates in an acoustic enclosure (AE) defined by two paraboloids forming a closed region (lens-shaped structure) with (I) rigid-wall boundary conditions, and (II) pressure-release boundaries.[3,4] In the first case with rigid walls (I), the normal surface velocity component vanishes on the boundary [$\nabla P \cdot \mathbf{n} = 0$, where P is the acoustic pressure and $\mathbf{n}$ is a surface normal vector (Neumann type)], whereas in the second case with pressure-release boundaries (II), the acoustic pressure must vanish on the boundary [$P = 0$ (Dirichlet type)]. The medium confined by the paraboloids is assumed to be air at room temperature. Both cases are solved quasianalytically based on the separability of the eigenfunction and the Frobenius power series assumption method.[2,5] For completeness, we note that the problem of an electron in a quantum dot of the same shape follows the same mathematics.[6]

In acoustics, the Helmholtz equation with rather general boundary conditions can be solved by means of, e.g., the finite-element method (FEM)[7–10] or the boundary-element method (BEM).[10–13] However, as FEM and BEM are numerical methods they cannot easily provide insight into the parametric dependencies of eigenmodes and eigenfrequencies as well as symmetry properties. Thus, it is preferable to employ analytical (and exact) techniques so as to obtain the eigenmodes and eigenfrequencies directly (including parameter dependencies and symmetry properties) whenever the problem domain geometry and boundary conditions are simple enough. This is, e.g., the case for cavities with rectangular, spherical, and cylindrical boundaries as is well-known from textbook examples.

In the present work, we solve the AE problem in PRC quasianalytically and determine eigenmodes and eigenvalues and their direct parameter dependence. A discussion of symmetry behaviors of the various eigenmodes is also given. The reason for studying the enclosure problem in PRC in acoustics is motivated by the need to understand better the sound field in lens-shaped geometries at intermediate frequencies (where the wavelength is comparable to the dimensions of the enclosure).[3,14] It is well-known that the geometrical acoustic/ray-tracing assumption as well as approximate low-frequency models (models where at least one dimension of the structure is much smaller than a wavelength, i.e., the acoustic pressure can be assumed to be constant along that dimension) fail to adequately describe the intermediate frequency range. In the example considered, the rigid-wall assumption and pressure-release boundary conditions are chosen, but the method described can be easily applied to the more general case with an effective wall acoustic impedance coefficient (mixed-type boundary conditions).[3,4]

II. THEORY

In this section, we show how to calculate the eigenfunctions and eigenvalues for the Helmholtz equation problem in a closed region of space defined by two paraboloids (Fig. 1). Such a surface can be described by two coordinates ξ and η which, together with a third coordinate ϕ, describe what is known as the parabolic rotational coordinates (PRC); the relationship to the Cartesian coordinates is as follows:

$$x = \xi\eta\cos\phi,$$
$$y = \xi\eta\sin\phi,$$
$$z = \tfrac{1}{2}(\eta^2 - \xi^2),$$
$$0 \leqslant \xi < \infty, \quad 0 \leqslant \eta < \infty, \quad 0 < \phi < 2\pi. \tag{1}$$

a)Electronic mail: Willatzen@mci.sdu.dk

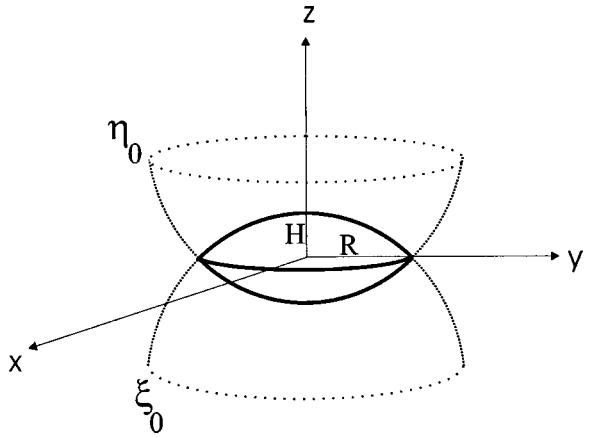


FIG. 1. Lens-shaped acoustic enclosure in parabolic rotational coordinates.

The constant ξ and η surfaces are paraboloids with respect to the positive z and negative z axes, respectively; e.g., a constant coordinate surface $\xi=\xi_0$ is the surface obtained by rotating a parabola about the positive z axis. Furthermore, it has a negative intercept on the z axis. $\xi=0$ ($\eta=0$) corresponds to the positive (negative) z axis. We note that the PRC system forms an orthogonal curvilinear coordinate system. Hence, the constant coordinate surfaces $\xi=\xi_0$ and $\eta=\eta_0$ (and also $\phi=\phi_0$) are orthogonal to each other. While $\xi=\xi_0$ and $\eta=\eta_0$ are used to define the closed region, more physically transparent parameters are the full height and radius (Fig. 1). The relationships of the various geometrical parameters are given as follows:

$$R=\xi_0\eta_0, \tag{2}$$

$$H=\frac{\xi_0^2+\eta_0^2}{2}, \tag{3}$$

$$V=\frac{\pi}{4}\xi_0^2\eta_0^2(\xi_0^2+\eta_0^2)=\pi R^2\frac{H}{2}, \tag{4}$$

where R, H, and V denote the radius, full height, and volume, respectively. The volume is obtained by integrating a volume element in parabolic rotational coordinates

$$V=\int_0^{2\pi}d\phi\int_0^{\eta_0}\int_0^{\xi_0}(\xi^2+\eta^2)\xi\eta\,d\eta\,d\xi.$$

Consider sound propagation in an AE defined by two paraboloids. The normal modes of the acoustic field are obtained by solving the wave equation

$$\nabla^2P+k^2P=0, \tag{5}$$

with P the acoustic pressure, $k=\omega/c$ the wave number, ω the angular frequency ($=2\pi f$), and c the sound speed in the medium confined by the two paraboloids subject to the boundary conditions (I)

$$\left.\frac{\partial P}{\partial\xi}\right|_{\xi=\xi_0}=0,\quad \left.\frac{\partial P}{\partial\eta}\right|_{\eta=\eta_0}=0, \tag{6}$$

corresponding to the rigid-wall case, or (II)

$$P(\xi=\xi_0)=0,\quad P(\eta=\eta_0)=0, \tag{7}$$

corresponding to the pressure-release boundary case.

The Helmholtz equation is known to be separable in parabolic rotational coordinates.[1,2] Writing

$$P(\xi,\eta,\phi)=M(\xi)N(\eta)\Phi(\phi), \tag{8}$$

one gets the following separated ordinary differential equations from Eq. (5):

$$\frac{d^2\Phi}{d\phi^2}+k_3^2\Phi=0, \tag{9}$$

$$\frac{1}{\xi}\frac{d}{d\xi}\left(\xi\frac{dM}{d\xi}\right)+\left[k_2+k^2\xi^2-\frac{k_3^2}{\xi^2}\right]M=0, \tag{10}$$

$$\frac{1}{\eta}\frac{d}{d\eta}\left(\eta\frac{dN}{d\eta}\right)-\left[k_2-k^2\eta^2+\frac{k_3^2}{\eta^2}\right]N=0, \tag{11}$$

where k_2 and k_3 are separation constants. The first equation, Eq. (9), is the familiar angular equation with solutions: $\exp(ik_3\phi)$. k_3 must be an integer for the solution to be single-valued (i.e., to return to its value after a rotation of 2π); hence, write $k_3=p$ (we will choose p positive since the $\pm p$ solutions are degenerate). The other two equations are known as the Bessel wave equations.[2] If $k=0$, one recovers the Bessel equations. We now consider how to solve Eqs. (10) and (11). First, let us rewrite the separation constants $k_2=q^2$ (while k_2 can be negative, we will provide an efficient method to account for it without additional work) and $k_3^2=p^2$. Then, Eq. (10) becomes

$$\frac{1}{\xi}\frac{d}{d\xi}\left(\xi\frac{dM}{d\xi}\right)+\left[q^2+k^2\xi^2-\frac{p^2}{\xi^2}\right]M=0. \tag{12}$$

Equation (12) can be solved using the Frobenius series method, a method also used by Moon and Spencer.[2]

The only singularity of Eq. (12) in the finite complex plane is a regular singular point at $\xi=0$. Expanding the solution about $\xi=0$, one writes

$$M(\xi)=\sum_{m=0}^{\infty}a_m\xi^{m+\sigma}, \tag{13}$$

and the coefficients a_m and the index σ are to be determined. The procedure is described in any mathematical physics book. For example, in Arfken,[5] it is shown that, if the differential equation is

$$x^2y''+xp(x)y'+q(x)y=0,$$

with

$$p(x)=\sum_{m=0}^{\infty}p_mx^m,\quad q(x)=\sum_{m=0}^{\infty}q_mx^m,$$

the indicial equation (which gives σ) is

$$\Theta(\sigma)\equiv\sigma(\sigma-1)+p_0\sigma+q_0=0,$$

and the recurrence relation (which gives the a_m's) is

$$a_n=-\frac{\sum_{m=1}^{n}a_{n-m}[(\sigma+n-m)p_m+q_m]}{\Theta(\sigma+n)}.$$

In our case, we have

$$\sigma = \pm p, \tag{14}$$

$$a_n = -\frac{q^2 a_{n-2} + k^2 a_{n-4}}{(p+n)^2 - p^2}. \tag{15}$$

Here, p is an integer; hence, the indicial equation only leads to one independent solution (will choose $\sigma = +p$ as the other solution diverges at $\xi = 0$). Choosing $a_0 = (q/2)^p/[\Gamma(p+1)]$, where Γ is the gamma function, one obtains the series expansion of the Bessel wave function[2]

$$M(\xi) = \frac{(q\xi/2)^p}{\Gamma(p+1)}\left\{1 - \frac{(q\xi/2)^2}{1!(p+1)} + \frac{(q\xi/2)^4}{2!(p+1)(p+2)} \times\left[1 - \frac{4(p+1)k^2}{q^4}\right] + \cdots\right\}. \tag{16}$$

This series solution is easily amenable to numerical work. Furthermore, note that Eq. (16) is completely real.

The solution to Eq. (11) is obtained in a similar fashion to $M(\xi)$. Indeed, the recurrence relation, Eq. (15), applies with the change $q^2 \to -q^2$ (and replacing a_n with b_n)

$$N(\eta) = \sum_{m=0}^{\infty} b_m \eta^{m+p}, \tag{17}$$

$$b_n = -\frac{-q^2 b_{n-2} + k^2 b_{n-4}}{(p+n)^2 - p^2}, \tag{18}$$

$$N(\eta) = \frac{(q\eta/2)^p}{\Gamma(p+1)}\left\{1 + \frac{(q\eta/2)^2}{1!(p+1)} + \frac{(q\eta/2)^4}{2!(p+1)(p+2)} \times\left[1 - \frac{4(p+1)k^2}{q^4}\right] + \cdots\right\}. \tag{19}$$

The boundary conditions, Eq. (7), now become

$$M(\xi = \xi_0) = \sum_{m=0}^{\infty} a_m \xi_0^{m+p} = 0,$$
$$N(\eta = \eta_0) = \sum_{m=0}^{\infty} b_m \eta_0^{m+p} = 0. \tag{20}$$

Similarly, the boundary conditions (6) become

$$\left.\frac{dM}{d\xi}\right|_{\xi=\xi_0} = \sum_{m=0}^{\infty} a_m (m+p)\xi_0^{m+p-1} = 0,$$
$$\left.\frac{dN}{d\eta}\right|_{\eta=\eta_0} = \sum_{m=0}^{\infty} b_m (m+p)\eta_0^{m+p-1} = 0. \tag{21}$$

As mentioned earlier, the coordinate pair (ξ_0, η_0) defines the closed region boundary.

Equations (20) and (21) are solved by first specifying p (an integer). For each p, there remains two unknowns, k (related to the frequency) and q, a separation constant. A simple procedure is to specify q and find the k values for each function M and N to have zeros; the boundary condition is satisfied when the two q (and k) values are identical.

It is possible for the separation constant k_2 in Eqs. (10) and (11) to be negative (that is, of the form $k_2 = -q^2$). Instead of rederiving all the results to account for this case explicitly, we make use of the fact that Eqs. (10) and (11) only differ in the sign of that same separation constant. Hence, given a (ξ_0, η_0) structure, one can find possible bound states for negative k_2 by instead finding the positive k_2 solutions of the (η_0, ξ_0) structure. This result reflects the fact that the (ξ_0, η_0) structure and the (η_0, ξ_0) structure are just mirror images of each other in the xy plane [they have the same radii, height, and volume as one can easily verify using Eqs. (2)–(4)]; hence, they have to have identical frequency spectra.

III. RESULTS AND DISCUSSIONS

In this section, numerical results of eigenfunctions and eigenvalues are given for structures defined by two paraboloids using parabolic rotational coordinates. The structures studied were a symmetrical AE (with $\xi_0 = \eta_0 = 1$ m) and asymmetrical structures (i.e., those having $\xi_0 \neq \eta_0$) with the same volume as the symmetrical one. They were all solved using the two boundary conditions discussed above.

A. Acoustic enclosures in PRC—rigid-wall boundary conditions (I)

Consider a symmetrical enclosure with rigid walls (defined by $\xi_0^2 = \eta_0^2 = 1$ m) containing air at room temperature (wave speed equal to $c = 343$ m/s). Eigenmodes are calculated by scanning in k for a given q (or q^2) for the intersection points (q^2, k) where the two functions $dM/d\xi|_{\xi=\xi_0}$ and $dN/d\eta|_{\eta=\eta_0}$ are zero simultaneously; see Fig. 2. The first plot [Fig. 2(a)] is for the case $p = 0$. Close inspection shows that $(k, q^2) = (0,0)$ is a solution to the rigid-wall problem. Indeed, for $q^2 = k = p = 0$, the ODEs [Eqs. (10)–(11)] are satisfied if M and N are constant functions. Thus, the acoustic pressure is constant (inside the enclosure) and the rigid-wall boundary conditions are trivially satisfied. Although $(q^2, k) = (0,0)$ is a solution when $p = 0$, this solution is uninteresting (known as the trivial solution) when discussing acoustic properties of the enclosure. In actual fact, trivial solutions of this kind (with eigenvalue $k = 0$) exist for *any* Helmholtz equation problem with Neumann boundary conditions imposed on the domain boundary for a general domain geometry. The first true solutions (for $p = 0$) are $(q^2, k) = (9.8171, 4.1341)$ m^{-1} and $(q^2, k) = (-9.8171, 4.1341)$ m^{-1} corresponding to an eigenfrequency of $f = kc/2\pi = 225.7$ Hz. In Fig. 2(b), the (q^2, k) plot of simultaneous zeros of M and N is shown for $p = 1$ for the symmetrical structure. In this case, the first solution in k is $(q^2, k) = (0, 2.331)$ m^{-1} corresponding to an eigenfrequency of $f = 127.2$ Hz. This is the lowest possible value for f amenable with the boundary conditions imposed. Thus, this frequency is the fundamental frequency of the AE. In actual fact, when $p = 2$, the lowest possible k value is 3.6824 m^{-1} ($f = 201$ Hz), and numerical results show that the lowest value of k increases as p increases (for p above 1!). In Fig. 2(c), the (q^2, k) plot of simultaneous zeros of M and N is shown for the case of an *asymmetrical* AE with surface parameters $\xi_0^2 = 1.2$ m and $\eta_0^2 = 0.8236$ m (having the same volume as the symmetrical AE with $\xi_0^2 = \eta_0^2 = 1$ m). The first solution

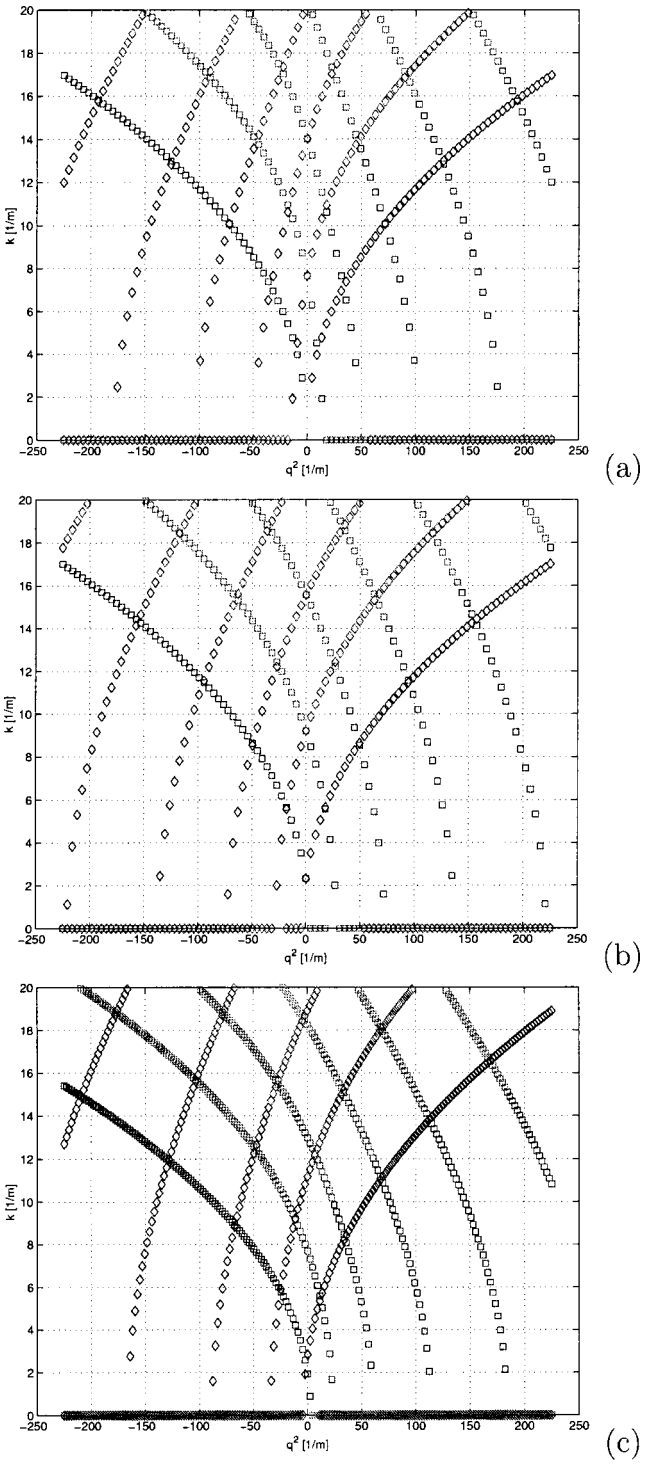


FIG. 2. Plot of k versus q^2 for three AEs with the same volume imposing rigid-wall boundary conditions. The three figures correspond to the cases: (a) $\xi_0^2=1$ m; $\eta_0^2=1$ m; $p=0$, and (b) $\xi_0^2=1$ m; $\eta_0^2=1$ m; $p=1$, and (c) $\xi_0^2=1.2$ m; $\eta_0^2=0.8236$ m; $p=1$, respectively. The AE volume is 1.57 m^3.

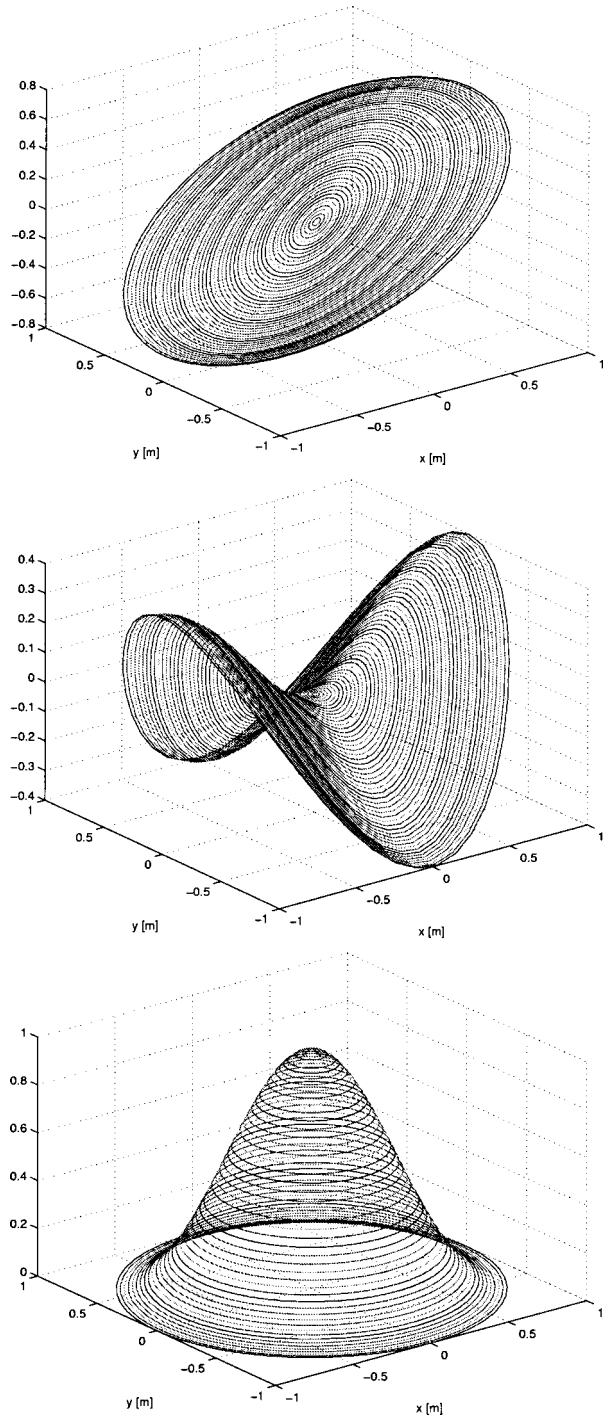


FIG. 3. Plot of the first three eigenmodes in the xy plane (for $z=0$) for the symmetrical AE with rigid walls. The first, second, and third modes are shown in the upper, middle, and lower plots, respectively.

$(q^2,k)=(-1.281,2.3093)$ m^{-1} is not symmetric with respect to a mirror reflection in the $z=0$ plane.

In Fig. 3(a), the fundamental mode for the symmetrical AE is plotted ($\xi_0^2=\eta_0^2=1$ m). We are plotting the pressure field P inside the lens-shaped AE in the $z=0$ plane as a contour plot. Evidently, the acoustic pressure gradient along the normal to the surface of the enclosure is zero but the pressure on the rim depends on the angle ϕ because this state corresponds to $p=1$ and $P\propto\exp(ip\phi)$. In Fig. 3(b), the second-lowest eigenmode corresponding to $p=2$ is plotted. Again, it is found that the acoustic pressure gradient along the normal to the AE surface is zero but the pressure changes (strongly) due to the $\exp(ip\phi)$ dependence. In Fig. 3(c), we show the first interesting eigenmode for $p=0$ corresponding to $(q^2,k)=(9.8171,4.1341)$ m^{-1}. This mode clearly satisfies the rigid-wall boundary conditions and is constant on the AE surface since $p=0$.

In Fig. 4(a), a plot of the fundamental mode frequency

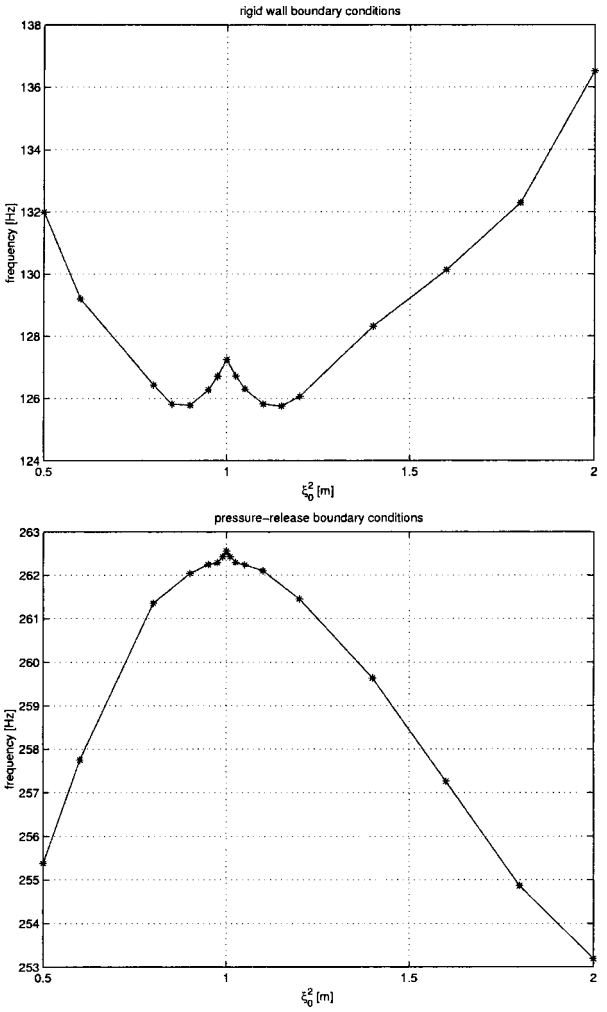


FIG. 4. Fundamental frequency for various AEs of the same volume. The top (bottom) plot corresponds to the rigid-wall (pressure-release boundary) case. The volume is the same as for Fig. 2.

as a function of the surface parameter ξ_0^2 is given for a series of AEs having the same volume equal to that of a symmetrical AE with $\xi_0^2=\eta_0^2=1$ m. In all cases, the p parameter value of the fundamental mode is 1. The symmetrical enclosure is characterized by a fundamental mode frequency above that of the asymmetrical enclosures in the range $\xi_0^2 \approx 0.9-1.1$ m. The explanation is that, when ξ_0^2 differs from 1 m, the enclosure becomes less flat along the z direction, being the smallest dimension of the enclosure. Thus, the fundamental mode frequency decreases with increasing asymmetry.

B. Acoustic enclosures in PRC—pressure-release boundaries (II)

Eigenmodes are now found by scanning in k for a given q (or q^2) for the intersection points (q^2,k) where the two functions $M(\xi=\xi_0)$ and $N(\eta=\eta_0)$ are zero simultaneously; see Fig. 5. The first plot [Fig. 5(a)] is for the case $p=0$. Opposite to the case with rigid walls, the parameter set $q^2=k=p=0$ is not a solution to the AE problem with pressure-release boundaries because M and N both being constants cannot satisfy the boundary conditions unless M and N are identically zero. In actual fact, the first intersection point occurs at $(q^2,k)=(0,4.809\,64)$ m^{-1} for $p=0$ corresponding to an eigenfrequency of 262.6 Hz, which is somewhat higher as compared to the 225.7 Hz found as the lowest eigenfrequency in the case of rigid walls (when $p=0$). Note also that the $(q^2,k)=(0,4.809\,64)$ m^{-1} indeed is the fundamental mode for pressure-release boundaries since numerical

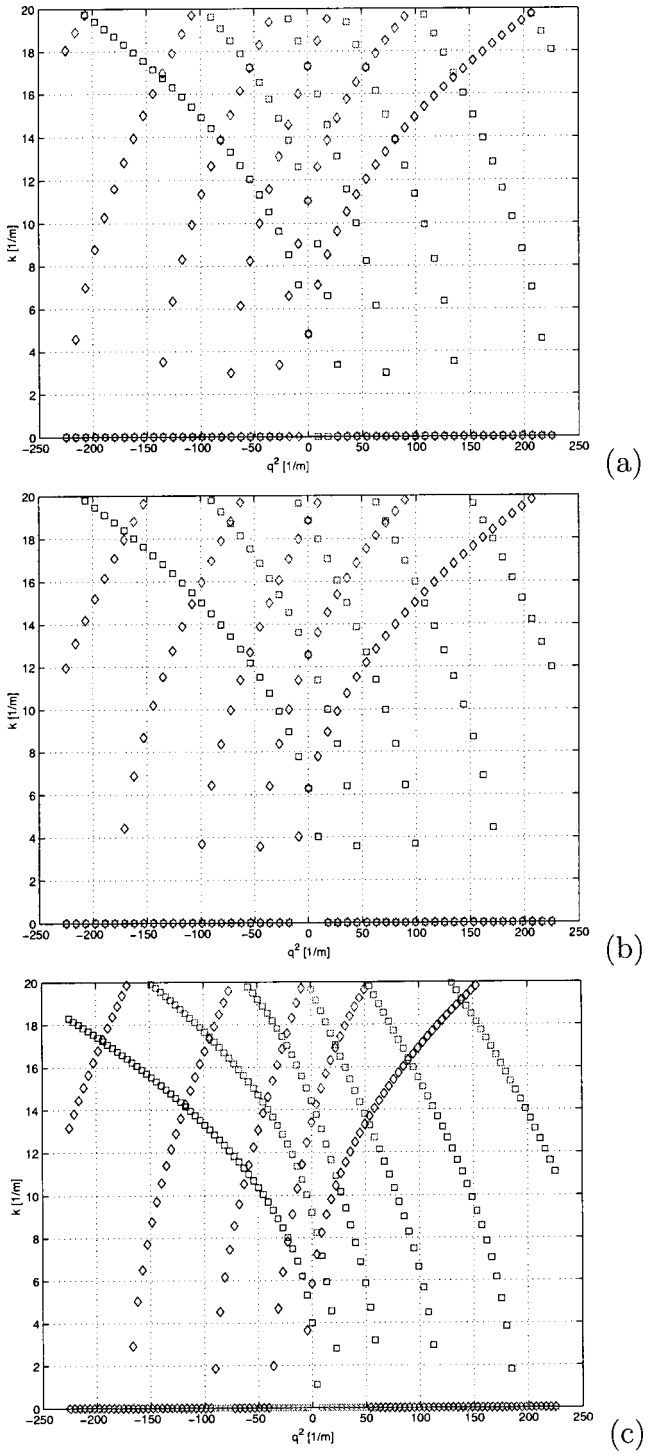


FIG. 5. Plot of k versus q^2 for three AEs with the same volume imposing pressure-release boundaries. The three figures correspond to the cases: (a) $\xi_0^2=1$ m; $\eta_0^2=1$ m; $p=0$, and (b) $\xi_0^2=1$ m; $\eta_0^2=1$ m; $p=1$, and (c) $\xi_0^2=1.2$ m; $\eta_0^2=0.8236$ m; $p=0$, respectively. The AE volume is 1.57 m^3.

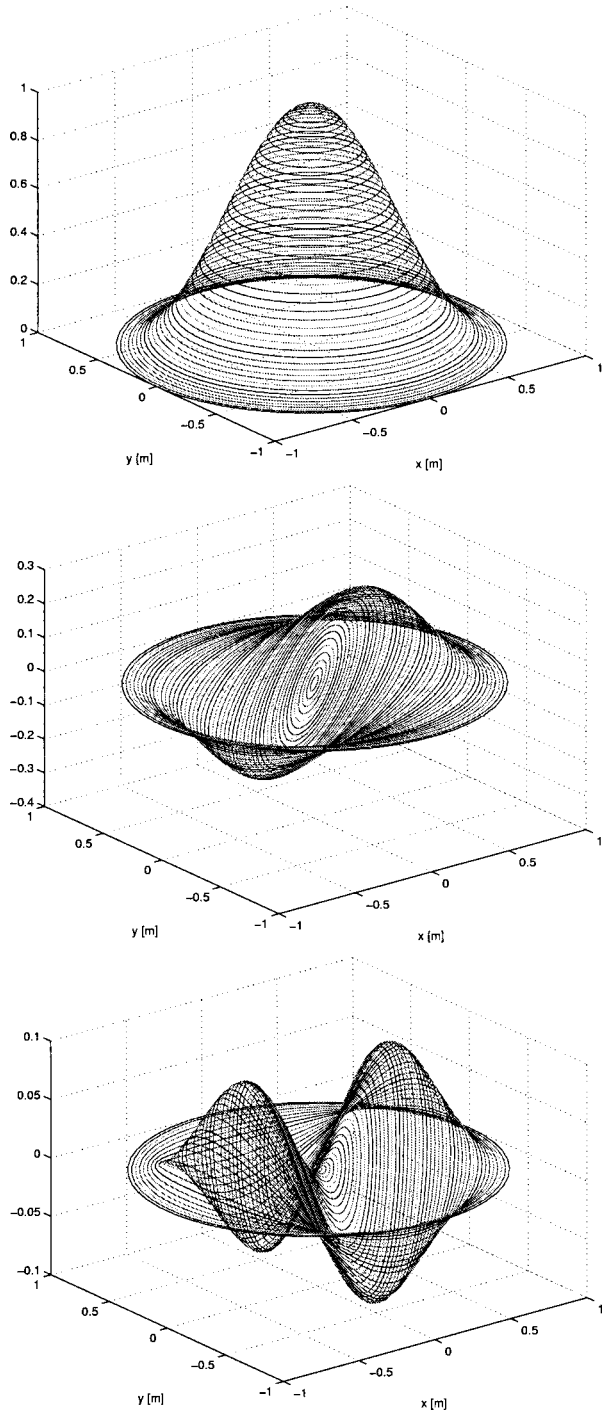


FIG. 6. Plot of the first three eigenmodes in the xy plane (for $z=0$) for the symmetrical AE with pressure-release boundaries. The first, second, and third modes are shown in the upper, middle, and lower plots, respectively.

results confirm that the lowest number of k values increases as p increases (for $p=0$ and above). In Fig. 5(b), a plot of intersection points where $M(\xi=\xi_0)$ and $N(\eta=\eta_0)$ are zero simultaneously is shown for the case $p=1$. Evidently, the first solution is $(q^2,k)=(0,6.2832)$ m^{-1} corresponding to an eigenfrequency of $f=343$ Hz. This is the first overtone for the symmetrical structure. In Fig. 5(c), intersection points where $M(\xi=\xi_0)$ and $N(\eta=\eta_0)$ are zero simultaneously are shown in (q^2,k) space for an asymmetrical AE with $\xi_0^2=0.8$ m (having the same volume as the symmetrical AE with $\xi_0^2=\eta_0^2=1$ m) and $p=0$. Similar to the case with rigid-wall boundary conditions, pair states associated with parameter values (q^2,k) and $(-q^2,k)$ are not degenerate as a consequence of the asymmetry. In actual fact, the fundamental mode eigenfrequency for this AE is $f=261.5$ Hz corresponding to $(q^2,k)=(-2.491,4.7894)$ m^{-1}

Next, in Fig. 6(a) we show the fundamental mode of the symmetrical AE with $\xi_0^2=\eta_0^2=1$ m and $p=0$. Since pressure-release boundary conditions are imposed, the acoustic pressure must vanish on the rim of the AE as Fig. 6(a) reveals. It is interesting to observe that this mode has a shape similar to the second overtone for the case with rigid walls (because both states are characterized by $p=q^2=0$). In Fig. 6(b), the first overtone of the symmetrical AE with pressure-release boundaries is shown. Clearly, the acoustic pressure is zero on the rim as required by the boundary conditions but the acoustic-pressure angular dependence is apparent inside the AE due to $p=1$. Finally, in Fig. 6(c) the second overtone for the symmetrical AE is depicted corresponding to $p=2$.

We also investigated the shape dependence of the fundamental mode frequency for a series of AEs with pressure-release boundaries all having the same volume [Fig. 4(b)]. The fundamental mode frequency decreases from 262.5 Hz when $\xi_0^2=1$ m to 255.5 Hz when $\xi_0^2=0.5$ m and 253.0 Hz for an AE with $\xi_0^2=2$ m. The results shown in Fig. 4(b) are substantially different from those shown in Fig. 4(a), i.e., the boundary conditions considerably affect the frequency *level* of the fundamental mode (being 262.5 Hz for pressure-release boundaries and 127.0 Hz for rigid walls of the symmetrical AE).

IV. CONCLUSIONS

The Helmholtz equation is solved exactly in parabolic rotational coordinates for a closed region defined by two paraboloids using separability of the eigenfunction and the Frobenius method. Two examples are considered and solved quasianalytically: the problem of acoustic cavity eigenmodes and eigenfrequencies for (I) an AE with rigid walls (Neumann type), and (II) an AE with pressure-release boundaries (Dirichlet type). Specific attention is given to the analysis of the shape dependence of the first eigenfrequency (fundamental mode) for both examples. It is shown that the same enclosure with different boundary conditions imposed has very different fundamental sound frequency values, and that the shape dependence of the spectrum is quite different for the two cases of boundary conditions analyzed.

[1] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[2] P. Moon and D. E. Spencer, *Field Theory Handbook* (Springer, Berlin, 1961).
[3] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[4] H. Kuttruff, *Room Acoustics*, 2nd ed. (Applied Science, London, 1979).
[5] G. Arfken, *Mathematical Methods for Physicists*, 3rd ed. (Academic, Orlando, 1985).
[6] L. C. Lew Yan Voon and M. Willatzen, "Confined states in lens-shaped quantum dots," J. Phys.: Condens. Matter **14**, 13667–13678 (2002).
[7] O. Z. Zienkiewicz, *The Finite Element Method* (McGraw-Hill, New York, 1977), Vols. I and II.
[8] A. Safjan, L. Demkowicz, and J. T. Oden, "Adaptive finite element meth-

ods for hyperbolic systems with applications to acoustics,'' Int. J. Numer. Methods Eng. **32**, 677–302 (1991).
[9] S. Kopuz and N. Lalor, ''Analysis of interior acoustic fields using the finite element method and the boundary element method,'' Appl. Acoust. **45**, 193–210 (1995).
[10] J. Raamachandran, *Boundary and Finite Elements: Theory and Practice* (Narosa, New Delhi, India, 1999).
[11] C. A. Brebbia and R. R. Ciskowski, *Boundary Element Methods in Acoustics* (Computational Mechanics, 1991).
[12] S. Kirkup, *The Boundary Element Method in Acoustics* (Integrated Sound Software, 1998).
[13] A. F. Seybert, C. Y. R. Cheng, and T. W. Wu, ''The solution of coupled interior/exterior acoustic problem using the boundary element method,'' J. Acoust. Soc. Am. **88**, 1612–168 (1990).
[14] *Handbook of Acoustics*, edited by M. J. Crocker (Wiley, New York, 1988).

Asymptotic solutions of the inhomogeneous Burgers equation

Michal Bednarik and Petr Konicek
Czech Technical University in Prague, FEE, Technicka 2, 166 27 Prague, Czech Republic

(Received 22 May 2003; accepted for publication 29 September 2003)

The method of the active second harmonic suppression in resonators is investigated in this paper both analytically and numerically. The resonator is driven by a piston which vibrates with two frequencies. The first one agrees with an eigenfrequency and the second one is equal to the two times higher eigenfrequency. The phase shift of the second piston motion is 180 deg. It is known that for this case it is possible to describe generation of the higher harmonics by means of the inhomogeneous Burgers equation. This model equation was solved for stationary state analytically by a number of authors but only for ideal fluids. Unlike their solutions, new asymptotic solutions are presented here which take into account dissipative effects. The asymptotic solutions are compared with numerical ones. For study of generation higher harmonics the solutions are developed in a spectral form. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1627839]

PACS numbers: 43.20.Ks, 43.25.Jh [MFH] Pages: 91–98

I. INTRODUCTION

Nowadays, we can observe the renewed interest in nonlinear standing waves. Theoretical and experimental studies have been made during the last decade. Scientific interest is stimulated by the possibility to accumulate acoustic energy in resonant cavities. However, the use of nonlinear standing waves is limited by the nonlinear attenuation that causes the acoustic saturation effects. The important characteristic of the resonators is the quality factor Q that shows how many times the amplitude of the steady-state wave is higher than the amplitude of vibration of the exciting piston. The Q factor depends on the amplitude of the vibrating piston due to the nonlinear attenuation. The nonlinear attenuation is connected with nonlinear acoustic wave interactions when we can observe generation of higher harmonics. As the thermoviscous attenuation is proportional to the square of frequency, it is possible to decrease the nonlinear attenuation by suppression of the wave cascade processes. It means that the energy transfer from the fundamental harmonic into higher ones is reduced, and for this reason the acoustic saturation effects are also suppressed. The suppression of acoustic saturation causes both the amplitude of the steady-state wave and the Q factor to increase. The higher Q factor means that more acoustic energy is accumulated in the resonators. The resonators of the high-Q factor are used for thermoacoustic engines, acoustic compressors, chemical disintegrating devices, etc. (see, e.g., Ref. 1).

There is quite a large number of methods enabling suppression of the nonlinear attenuation and thus an increase in the quality factor of the given resonator. One of the methods is called macrosonic resonance synthesis (see Refs. 2–5), which uses resonators of variable diameters. These resonators have eigenfrequencies which are nonequidistant, that is, the higher eigenfrequencies are not integer multiples of the fundamental one. If the source frequency is equal to the fundametal eigenfrequency of the resonator, then the generated higher harmonics does not coincide with the higher eigenfrequencies. It means that the resonator eigenfrequencies interact with higher harmonics ineffectively, and thus the harmonics does not achieve levels like we can observe for the resonators with a constant diameter.

When we want to control generation of higher harmonics, it is possible to place a selective absorber directly into the resonator. If the selective absorber is used for the second harmonic we can interrupt the cascade processes, see, e.g., Refs. 6 and 7.

It is known (e.g., Ref. 8) that the nonlinear attenuation due to the higher harmonics may be avoided by introducing phase-speed dispersion as well. This fact was utilized by Sugimoto in Ref. 9. In this work he used a periodic array of Helmholtz resonators to obtain the artificial dispersion.

Another promising method is based on the active suppression of the second harmonic component of the acoustic wave.[10–12] For this purpose it is necessary to employ the piston which vibrates with two frequencies. The first one agrees with an eigenfrequency f and the second one is equal to the eigenfrequency $2f$. The phase shift of the second piston motion is 180 deg.

Many studies are only dedicated to the description of behavior of nonlinear standing waves in resonators without stressing suppression of the nonlinear attenuation; see Refs. 13–18. Wave processes in the resonators were studied experimentally in Refs. 19 and 20.

This paper deals with the method of the active second harmonic suppression mentioned above both analytically and numerically. It is known that for this case it is possible to describe generation of the higher harmonics by means of the inhomogeneous Burgers equation. The resonator is driven by a piston whose motion is characterized by two superposed sinusoidal motions. This problem was treated for stationary state in Refs. 11 and 12. However, the authors of these papers took into account only inviscid solutions. It means that discontinuities were contained in their solutions. Unlike these solutions, we present new asymptotic solutions which take into account dissipative effects. Some of the solutions are also presented in the spectral form that is more suitable for study of generation higher harmonics in the resonators. The asymptotic solutions are compared with numerical ones.

II. ASYMPTOTIC SOLUTION OF THE MODEL EQUATION

When describing the nonlinear plane standing waves in resonator of a constant radius, it is possible to start with the Kuznetsov's model equation for velocity potential ϕ in the second approximation (see, e.g., Ref. 21)

$$\frac{\partial^2\phi}{\partial t^2}-c_0^2\frac{\partial^2\phi}{\partial x^2}+\frac{\gamma-1}{2c_0^2}\frac{\partial}{\partial t}\left(\frac{\partial\phi}{\partial t}\right)^2+\frac{\partial}{\partial t}\left(\frac{\partial\phi}{\partial x}\right)^2=\frac{b}{\rho_0 c_0^2}\frac{\partial^3\phi}{\partial t^3}, \tag{1}$$

where x is the space coordinate in the direction of the resonator axis, t is time, c_0 is the small signal sound speed, ρ_0 is the ambient density of the fluid, $\gamma=c_p/c_V$ is Poisson's number, and c_p, or c_v is the specific heat under constant pressure, or volume, $b=\rho_0\delta$, where δ is the diffusivity of sound (see, e.g., Ref. 22).

In this resonator we can imagine the sound field as a superposition of simple waves propagating in opposite directions which are assumed not to interact in the volume of the resonator, and they are coupled only by the conditions on the walls of resonator; see Ref. 11. The next possible simplification is when we neglect the fact that the driving piston is moving and thus the position of the boundary of the resonator is unvarying with the time. This assumption is acceptable for very small amplitude of driving piston. With the abovementioned suppositions, we can find the solution of this equation in the following form:

$$\phi=\left[\mu\phi_+\left(\mu x,\mu t,\tau_+=t-\frac{x}{c_0}\right)-\mu\phi_-\left(\mu x,\mu t,\tau_-=t+\frac{x}{c_0}\right)\right], \tag{2}$$

where μ is a small parameter.

Substituting the expression (2) into Eq. (1) and neglecting the terms of the order three and higher, and supposing that counterpropagating waves do not interact we can get, likewise Ref. 15, the following two equations:

$$c_0\frac{\partial v_+}{\partial x}+\frac{\partial v_+}{\partial t}-\frac{\beta}{c_0}v_+\frac{\partial v_+}{\partial\tau_+}=\frac{b}{2\rho_0 c_0^2}\frac{\partial^2 v_+}{\partial\tau_+^2}, \tag{3}$$

$$-c_0\frac{\partial v_-}{\partial x}+\frac{\partial v_-}{\partial t}-\frac{\beta}{c_0}v_-\frac{\partial v_-}{\partial\tau_-}=\frac{b}{2\rho_0 c_0^2}\frac{\partial^2 v_-}{\partial\tau_-^2}. \tag{4}$$

We can write for an acoustic velocity

$$v=v_+-v_-, \tag{5}$$

where v_+ and v_- are solutions of Eqs. (3) and (4).

The length of the resonator of a constant diameter is labeled by L. It is valid for angular eigenfrequencies ω_n that

$$\omega_n=\frac{n\pi c_0}{L},\quad n=1,2,3,\ldots. \tag{6}$$

In the case that we consider the harmonic excitation of the standing waves with the piston at the position $x=L$, we can express the boundary and initial conditions as follows:

$$v=(v_+-v_-)_{x=0}=0, \tag{7}$$

$$v=(v_+-v_-)_{x=L}=v_{m1}\sin(\omega t)+v_{m2}\sin(2\omega t+\varphi), \tag{8}$$

and

$$v_\pm(t=0)=0, \tag{9}$$

where v_{m1} and v_{m2} are acoustic velocity amplitudes of the piston and φ is a phase shift. We assume that a piston vibrates with the angular frequency ω, which is equal to the $(2n+1)$-th eigenfrequency of the given resonator; it means that $\omega=\omega_{2n+1}$. This assumption causes higher harmonic components of an acoustic velocity to be in coincidence with eigenfrequencies. Equations (3) and (4) together with conditions (7), (8), and (9) can be solved by the method of successive approximation; see Ref. 15. On the basis of this method we obtain these model equations:

$$\frac{\partial\overline{v}_\pm}{\partial t}-\frac{\beta}{c_0}\overline{v}_\pm\frac{\partial\overline{v}_\pm}{\partial\tau_\pm}-\frac{b}{2\rho_0 c_0^2}\frac{\partial^2\overline{v}_\pm}{\partial\tau_\pm^2}=\frac{v_{m1}c_0}{2L}\sin(\omega\tau_\pm)+\frac{v_{m2}c_0}{2L}\sin(2\omega\tau_\pm+\varphi). \tag{10}$$

Equation (10) represent the inhomogeneous Burgers equation, where

$$\overline{v}_\pm(t,\tau_\pm)=v_\pm(t,\tau_\pm)\pm\frac{v_{m1}x}{2L}\sin(\omega\tau_\pm)\pm\frac{v_{m2}x}{2L}\sin(2\omega\tau_\pm+\varphi). \tag{11}$$

Substituting from Eqs. (11) to Eq. (5), we obtain

$$v(t,x)=\overline{v}_+-\overline{v}_--\frac{v_{m1}x}{L}\cos\left(\frac{\omega x}{c_0}\right)\sin(\omega t)-\frac{v_{m2}x}{L}\cos\left(\frac{2\omega x}{c_0}\right)\sin(2\omega t+\varphi). \tag{12}$$

It is more suitable to express Eqs. (10) in the dimensionless form

$$\frac{\partial\overline{V}_\pm}{\partial s}-\overline{V}_\pm\frac{\partial\overline{V}_\pm}{\partial y_\pm}-\frac{1}{\Gamma}\frac{\partial^2\overline{V}_\pm}{\partial y_\pm^2}=\sin(y_\pm)+p\sin(2y_\pm+\varphi), \tag{13}$$

where

$$s=\frac{t}{t_s},\quad \overline{V}_\pm=\frac{\overline{v}_\pm}{v_0},\quad y_\pm=\omega\tau_\pm,\quad \Gamma=\frac{2\rho_0 c_0\beta v_0}{b\omega},\quad p=\frac{v_{m2}}{v_{m1}},\quad v_0=\sqrt{\frac{v_{m1}c_0}{2\pi\beta}},\quad t_s=\frac{c_0}{\beta\omega v_0}. \tag{14}$$

Equations (13) have the same form for both counterpropagating waves, and consequently we can rewrite them for clarity as

$$\frac{\partial V}{\partial s}-V\frac{\partial V}{\partial y}-\frac{1}{\Gamma}\frac{\partial^2 V}{\partial y^2}=\sin(y)+p\sin(2y+\varphi). \tag{15}$$

Supposing the stationary state ($\partial V/\partial s=0$) and using the Cole–Hopf transformation (see, e.g., Refs. 8, 22)

$$V=\frac{2}{\Gamma}\frac{U'}{U}, \tag{16}$$

we obtain the following linear differential equation from Eq. (15):

$$U''+\left[a-\frac{\Gamma^2}{2}\cos(y)-\frac{\Gamma^2 p}{4}\cos(2y+\varphi)\right]U=0, \tag{17}$$

where the comma represents the derivative with respect to y. Equation (17) is called the Whittaker–Hill equation (see Refs. 23 and 24).

The symbol a in Eq. (17) is the integration constant which represents the eigenvalue of the Whittaker–Hill equation. Let Γ be large; then, we assume a solution of Eq. (17) in the form (see, e.g., Refs. 25 and 24)

$$U(y)=\exp[\Gamma g(y)]\Psi(y)[1+h_1(y)\Gamma^{-1}+h_2(y)\Gamma^{-2}+\cdots], \tag{18}$$

with

$$a=\alpha\Gamma^2+\beta\Gamma+q_0+q_1\Gamma^{-1}+q_2\Gamma^{-2}+\cdots. \tag{19}$$

For very large Γ, we can write

$$U\simeq\exp[\Gamma g(y)]\Psi(y), \tag{20}$$

and

$$a\simeq\alpha\Gamma^2+\beta\Gamma. \tag{21}$$

Substituting Eqs. (20) and (21) into Eq. (17), we get

$$\left[\Gamma^2(g')^2+\Gamma g''+\frac{2\Gamma}{\Psi}g'\Psi'+\alpha\Gamma^2+\beta\Gamma-\frac{\Gamma^2}{2}\cos(y)-\frac{\Gamma^2 p}{4}\cos(2y+\varphi)\right]\exp(\Gamma g)\Psi+\exp(\Gamma g)=0. \tag{22}$$

Comparing coefficients Γ^2 we get the equation

$$\left[(g')^2+\alpha-\frac{1}{2}\cos(y)-\frac{p}{4}\cos(2y+\varphi)\right]\exp(\Gamma g)\Psi=0. \tag{23}$$

We can arrange Eq. (23) into the form

$$(g')^2=-\alpha+\cos^2\left(\frac{y}{2}\right)+\frac{p}{2}\cos^2\left(y+\frac{\varphi}{2}\right)-\frac{1}{2}-\frac{p}{4}. \tag{24}$$

It is reasonable to find a value α making $g(y)$ periodic. This leads to

$$\alpha=-\frac{1}{2}-\frac{p}{4}. \tag{25}$$

Then, we can write

$$(g')^2=\cos^2\left(\frac{y}{2}\right)+\frac{p}{2}\cos^2\left(y+\frac{\varphi}{2}\right). \tag{26}$$

According to Ref. 11, we assume $\varphi=\pi$. Then, Eq. (26) can be written as

$$g'_{\pm}=\pm\cos\left(\frac{y}{2}\right)\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)}. \tag{27}$$

Solving Eq. (27), we have

$$g_{\pm}=\pm\sin\left(\frac{y}{2}\right)\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)}\pm\frac{\ln\left[\sqrt{2p}\sin\left(\frac{y}{2}\right)+\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)}\right]}{\sqrt{2p}}+A_{\pm}, \tag{28}$$

where $A_{\pm}$ are the integration constants which are equal to 0 for generation of standing waves in a resonance.

Now, comparing coefficients Γ in Eq. (22) we get the equation

$$\left(g''+\frac{2}{\Psi}g'\Psi'+\beta\right)\exp(\Gamma g)=0. \tag{29}$$

After arranging, Eq. (29) becomes

$$\Psi g''+2g'\Psi'+\beta\Psi=0. \tag{30}$$

The first and second derivative of the formula (28) take the following form:

$$g'_{\pm}=\pm\cos\left(\frac{y}{2}\right)\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)}. \tag{31}$$

$$g''_{\pm}=\mp\frac{\sin(y/2)[1-2p\cos(y)]}{2\sqrt{1+2p\sin^2(y/2)}}. \tag{32}$$

When $p\to 0$ Eq. (17) degenerates to the Mathieu's equation, which has the standard form for $f(z)$ (see, e.g., Ref. 25)

$$\frac{d^2 f}{dz^2}+[a-2q\cos(2z)]f=0. \tag{33}$$

For $p\to 0$ we can simplify Eqs. (28), (31), and (32)

$$g_{\pm}=2\sin\left(\frac{y}{2}\right), \tag{34}$$

$$g'_{\pm}=\pm\cos\left(\frac{y}{2}\right), \tag{35}$$

$$g''_{\pm}=\mp\frac{1}{2}\sin\left(\frac{y}{2}\right). \tag{36}$$

Substituting (35) and (36) into (30), we get

$$\Psi'_{\pm}+\left[\frac{\pm\beta}{2\cos\left(\frac{y}{2}\right)}-\frac{\sin\left(\frac{y}{2}\right)}{4\cos\left(\frac{y}{2}\right)}\right]\Psi_{\pm}=0. \tag{37}$$

Solving this equation we have

$$\Psi_{\pm}=\frac{C_{\pm}\left[\frac{1}{\cos(y/2)}+\tan\left(\frac{y}{2}\right)\right]^{\mp\beta}}{\sqrt{\cos\left(\frac{y}{2}\right)}}, \tag{38}$$

where C_+ and C_- are integrating constants. We can arrange Eq. (38) in this form

$$\Psi_+ = \frac{C_+\sqrt{2}\cos\left(\frac{y}{4}+\frac{\pi}{4}\right)}{\cos\left(\frac{y}{2}\right)}, \tag{39}$$

$$\Psi_- = \frac{C_-\sqrt{2}\sin\left(\frac{y}{4}+\frac{\pi}{4}\right)}{\cos\left(\frac{y}{2}\right)}, \tag{40}$$

where $\beta = 1/2$; see, e.g., Ref. 25. Substituting formulas (39), (40), and (34) into Eq. (20), we obtain

$$U_+(y) = \frac{C_+\sqrt{2}}{\cos\left(\frac{y}{2}\right)}\left\{\exp\left[\Gamma\sin\left(\frac{y}{2}\right)\right]\cos\left(\frac{y}{4}+\frac{\pi}{4}\right)\right\}, \tag{41}$$

$$U_-(y) = \frac{C_-\sqrt{2}}{\cos\left(\frac{y}{2}\right)}\left\{\exp\left[-\Gamma\sin\left(\frac{y}{2}\right)\right]\sin\left(\frac{y}{4}+\frac{\pi}{4}\right)\right\}. \tag{42}$$

Writing $-y$ for y in Eq. (42) gives (41), and vice versa.

We can write the solution as

$$U(y) \simeq C_+\frac{\sqrt{2}}{\cos\left(\frac{y}{2}\right)}\left\{\exp\left[\Gamma\sin\left(\frac{y}{2}\right)\right]\cos\left(\frac{y}{4}+\frac{\pi}{4}\right)\right\} + C_-\frac{\sqrt{2}}{\cos\left(\frac{y}{2}\right)}\left\{\exp\left[-\Gamma\sin\left(\frac{y}{2}\right)\right]\sin\left(\frac{y}{4}+\frac{\pi}{4}\right)\right\}. \tag{43}$$

Bearing in mind the interchangeability of Eq. (41) [Eq. (42) as mentioned above], we choose $C_+ = C_- = C/\sqrt{2}$. As the solution has to lead to the even Mathieu function ce_0 (see, e.g., Refs. 16 and 25), we can write

$$U(y) \simeq \frac{C}{\cos\left(\frac{y}{2}\right)}\left\{\exp\left[\Gamma\sin\left(\frac{y}{2}\right)\right]\cos\left(\frac{y}{4}+\frac{\pi}{4}\right) + \exp\left[-\Gamma\sin\left(\frac{y}{2}\right)\right]\sin\left(\frac{y}{4}+\frac{\pi}{4}\right)\right\} \simeq \mathrm{ce}_0\left(\frac{y}{2},\frac{\Gamma^2}{4}\right), \tag{44}$$

for the open interval $-\pi < y < \pi$.

Assuming that $p \to \infty$ we can continue likewise for $p \to 0$. For this case Eq. (17) degenerates to the Mathieu's equation again. It is possible to simplify Eqs. (31) and (32)

$$g'_\pm \simeq \pm\frac{\sqrt{2p}}{2}\sin(y), \tag{45}$$

$$g''_\pm \simeq \pm\frac{\sqrt{2p}}{2}\cos(y). \tag{46}$$

Putting Eqs. (45) and (46) into Eq. (30), we get this equation

$$\Psi'_\pm + \frac{\cos(y)}{2\sin(y)}\Psi_\pm \pm \frac{\beta}{\sqrt{2p}\sin(y)}\Psi_\pm = 0. \tag{47}$$

Solving Eq. (47), we obtain

$$\Psi_\pm = K_\pm\frac{\left[\frac{1-\cos(y)}{\sin(y)}\right]^{\mp\beta/\sqrt{2p}}}{\sqrt{\sin(y)}}, \tag{48}$$

where $K_\pm$ are integrating constants. Using the following equalities:

$$\sin(y) = \cos(y-\pi/2), \tag{49}$$

$$\cos(y) = -\sin(y-\pi/2), \tag{50}$$

we can express Eq. (48) as

$$\Psi_\pm = \frac{K_\pm\left[\frac{1}{\cos(y-\pi/2)}+\tan(y-\pi/2)\right]^{\mp\beta/\sqrt{2p}}}{\sqrt{\cos(y-\pi/2)}}. \tag{51}$$

Comparing the expression (51) with the expression (38) we can see that $\beta = \sqrt{2p}/2$.

Substituting formulas (28) and (48) into Eq. (20), we get for the open interval $-\pi < y < \pi$

$$U(y) \simeq K\{\exp[f(y)]\Psi_+(y) + \exp[-f(y)]\Psi_-(y)\} \simeq \mathrm{ce}_0\left(y-\frac{\pi}{2},\frac{\Gamma^2 p}{8}\right), \tag{52}$$

where

$$f(y) = \Gamma\sin\left(\frac{y}{2}\right)\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)} + \frac{\Gamma\ln\left[\sqrt{2p}\sin\left(\frac{y}{2}\right)+\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)}\right]}{\sqrt{2p}}. \tag{53}$$

Using the expression (16), we can write

$$V = 2\frac{[\xi(y;p)\Psi_+ + \Psi'_+]\exp(f)}{\exp(f)\Psi_+ + \exp(-f)\Psi_-} - 2\frac{[\xi(y;p)\Psi_- - \Psi'_-]\exp(-f)}{\exp(f)\Psi_+ + \exp(-f)\Psi_-}, \tag{54}$$

where

$$\xi(y;p) = \cos\left(\frac{y}{2}\right)\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)}. \tag{55}$$

Reminding that $\Gamma \gg 1$ we can simplify the solution (54)

$$V(y)\simeq 2\cos\left(\frac{y}{2}\right)\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)}\times\tanh\left\{\Gamma\sin\left(\frac{y}{2}\right)\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)}+\Gamma\frac{\ln\left[\sqrt{2p}\sin\left(\frac{y}{2}\right)+\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)}\right]}{\sqrt{2p}}\right\},\tag{56}$$

for the closed interval $-\pi\leqslant y\leqslant\pi$.

We can dare this simplification because the functions $\Psi_{\pm}$ behave "sensibly" for $p\to 0$ and $p\to\infty$. Below we can see how it is possible to generalize the "sensible" behavior of the functions $\Psi_{\pm}$ for all $p\geqslant 0$.

The solution (56) represents the approximation solution of the inhomogeneous Burgers equation (15). This can be reduced to the known solutions when we take into account the following two cases:

When Γ tends to infinity (an ideal fluid) we get from Eq. (56)

$$V(y)=2\cos\left(\frac{y}{2}\right)\mathrm{sign}(y)\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)},\tag{57}$$

where sign represents the function signum. The solution (57) was presented in Refs. 11 or 26.

When $p\to 0$, then the solution (56) and (57) takes the form (see Ref. 13)

$$V(y)\simeq 2\cos\left(\frac{y}{2}\right)\tanh\left[2\Gamma\sin\left(\frac{y}{2}\right)\right],\tag{58}$$

and

$$V(y)=2\cos\left(\frac{y}{2}\right)\mathrm{sign}(y).\tag{59}$$

The results from the asymptotic solution, which is presented in Ref. 27, are comparable with results from the solution (58). A similar solution to the solution (58) can be obtained by matching the finite-width shock front to the solution (59) ($\Gamma\to\infty$) as well. The expression (58) represents the approximate solution of the inhomogeneous Burgers equation (see, e.g., Refs. 16 and 14)

$$\frac{\partial V}{\partial s}-V\frac{\partial V}{\partial y}-\frac{1}{\Gamma}\frac{\partial^2 V}{\partial y^2}=\sin(y).\tag{60}$$

When p is very large the second term in the argument of the hyperbolic tangent of the solution (56) is insignificant in comparison with the first term, and then we can simplify the solution (56) to the form

$$V(y)\simeq 2\cos\left(\frac{y}{2}\right)\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)}\times\tanh\left[\Gamma\sin\left(\frac{y}{2}\right)\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)}\right].\tag{61}$$

As p is not very large, both terms in the argument of the hyperbolic tangent of the solution (56) are comparable; thus, we can simplify the solution (56)

$$V(y)\simeq 2\cos\left(\frac{y}{2}\right)\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)}\times\tanh\left[2\Gamma\sin\left(\frac{y}{2}\right)\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)}\right].\tag{62}$$

When p is very large we can reduce the solution (62) to

$$V(y)\simeq\sqrt{2p}\sin(y).\tag{63}$$

We can express the solution (62) by means of the Fourier series

$$V\simeq\sum_{n=1}^{\infty}V_n\sin(ny),\tag{64}$$

where

$$V_n=\frac{2}{\pi}\int_0^{\pi}2\cos\left(\frac{y}{2}\right)\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)}\times\tanh\left[2\Gamma\sin\left(\frac{y}{2}\right)\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)}\right]\sin(ny)\mathrm{d}y.\tag{65}$$

When $y\in(-\varepsilon;\varepsilon)$, where ε is small, it is possible to put $\sin(y/2)\simeq y/2$. Then, we can replace the integrated function (62) for $p\leqslant 0.5$ as

$$\begin{aligned}&2\cos\left(\frac{y}{2}\right)\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)}\times\tanh\left[2\Gamma\sin\left(\frac{y}{2}\right)\sqrt{1+2p\sin^2\left(\frac{y}{2}\right)}\right]\sin(ny)\\&\approx 2\cos\left(\frac{y}{2}\right)\sqrt{1+p\sin^2\left(\frac{y}{2}\right)}\times\tanh\left[\Gamma y\sqrt{1+p\sin^2\left(\frac{y}{2}\right)}\right]\sin(ny),\quad -\pi\leqslant y\leqslant\pi.\end{aligned}\tag{66}$$

After replacing the integrated function with (66) in Eq. (65) and differentiating with respect to Γ, we obtain

$$\frac{\partial V_n}{\partial\Gamma}\simeq\frac{4}{\pi}\int_0^{\pi}\frac{y\cos\left(\frac{y}{2}\right)\left[1+p\sin^2\left(\frac{y}{2}\right)\right]\sin(ny)}{\cosh^2\left[\Gamma y\sqrt{1+p\sin^2\left(\frac{y}{2}\right)}\right]}\mathrm{d}y.\tag{67}$$

For $p\leqslant 0.5$ and $\Gamma\gg 1$ we can simplify the argument of the hyperbolic cosine in Eq. (67)

$$\frac{\partial V_n}{\partial\Gamma}\simeq\frac{4}{\pi}\int_0^{\pi}\frac{y\cos\left(\frac{y}{2}\right)\left[1+p\sin^2\left(\frac{y}{2}\right)\right]\sin(ny)}{\cosh^2(\Gamma y)}\mathrm{d}y.\tag{68}$$

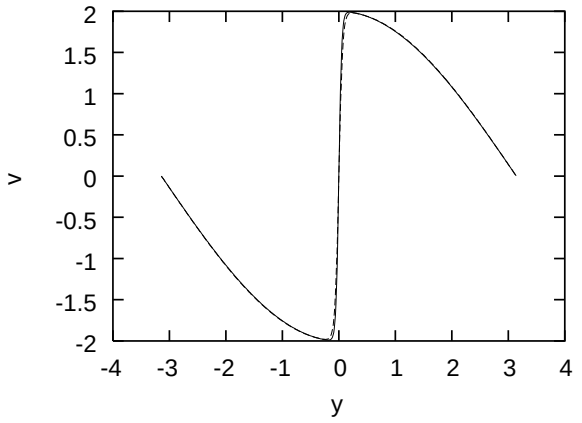


FIG. 1. Comparison of the asymptotic solution (solid line) and the numerical solution (dashed line), $\Gamma=20$, $p=0$.

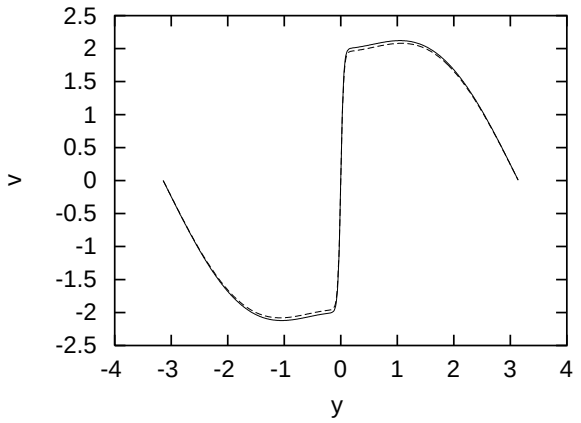


FIG. 2. Comparison of the asymptotic solution (solid line) and the numerical solution (dashed line), $\Gamma=20$, $p=1$.

As the integrated function in Eq. (68) has significant values only in the interval $0\leqslant y\leqslant \pi$, we can replace the upper limit by ∞

$$\frac{\partial V_n}{\partial \Gamma}\simeq \frac{4}{\pi}\int_0^{\infty}\frac{y\cos\left(\frac{y}{2}\right)\left[1+p\sin^2\left(\frac{y}{2}\right)\right]\sin(ny)}{\cosh^2(\Gamma y)}\,\mathrm{d}y. \quad (69)$$

Further, we can arrange the numerator of the integrated function in Eq. (69) as

$$\frac{\partial V_n}{\partial \Gamma}\simeq \frac{2+\frac{p}{2}}{\pi}\int_0^{\infty}\frac{y\left\{\sin\left[\left(n+\frac{1}{2}\right)y\right]+\sin\left[\left(n-\frac{1}{2}\right)y\right]\right\}}{\cosh^2(\Gamma y)}\,\mathrm{d}y - \frac{\frac{p}{2}}{\pi}\int_0^{\infty}\frac{y\left\{\sin\left[\left(n+\frac{3}{2}\right)y\right]+\sin\left[\left(n-\frac{3}{2}\right)y\right]\right\}}{\cosh^2(\Gamma y)}\,\mathrm{d}y. \quad (70)$$

After integrating, we have

$$\frac{\partial V_n}{\partial \Gamma}\simeq \frac{2+\frac{p}{2}}{2}\frac{\partial}{\partial\Gamma}\left\{\frac{1}{\Gamma\sinh\left[\frac{\pi(n+1/2)}{2\Gamma}\right]}\right\} + \frac{2+\frac{p}{2}}{2}\frac{\partial}{\partial\Gamma}\left\{\frac{1}{\Gamma\sinh\left[\frac{\pi(n-1/2)}{2\Gamma}\right]}\right\} - \frac{p}{4}\frac{\partial}{\partial\Gamma}\left\{\frac{1}{\Gamma\sinh\left[\frac{\pi(n+3/2)}{2\Gamma}\right]}\right\} - \frac{p}{4}\frac{\partial}{\partial\Gamma}\left\{\frac{1}{\Gamma\sinh\left[\frac{\pi(n-3/2)}{2\Gamma}\right]}\right\}. \quad (71)$$

Then, we can write

$$V_n\simeq \frac{2+\frac{p}{2}}{2\Gamma\sinh\left[\frac{\pi(n+1/2)}{2\Gamma}\right]}+\frac{2+\frac{p}{2}}{2\Gamma\sinh\left[\frac{\pi(n-1/2)}{2\Gamma}\right]} - \frac{p}{4\Gamma\sinh\left[\frac{\pi(n+3/2)}{2\Gamma}\right]}-\frac{p}{4\Gamma\sinh\left[\frac{\pi(n-3/2)}{2\Gamma}\right]}. \quad (72)$$

When $\Gamma\to\infty$ (ideal fluid), we obtain from (72) this expression:

$$V_n\simeq -\frac{16n(9+2p-4n^2)}{\pi(16n^4-40n^2+9)}. \quad (73)$$

When $p=0$, then Eq. (73) takes the form

$$V_n=\frac{16n}{\pi(4n^2-1)}. \quad (74)$$

With the help of expressions (12) and (64), we can write

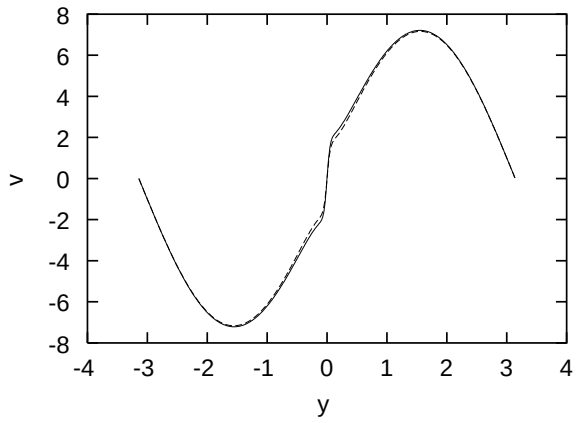


FIG. 3. Comparison of the asymptotic solution (solid line) and the numerical solution (dashed line), $\Gamma=20$, $p=25$.

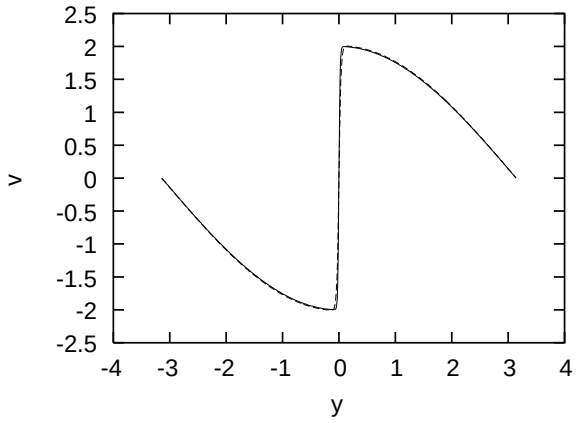


FIG. 4. Comparison of the asymptotic solution (solid line) and the numerical solution (dashed line), $\Gamma=50$, $p=0$.

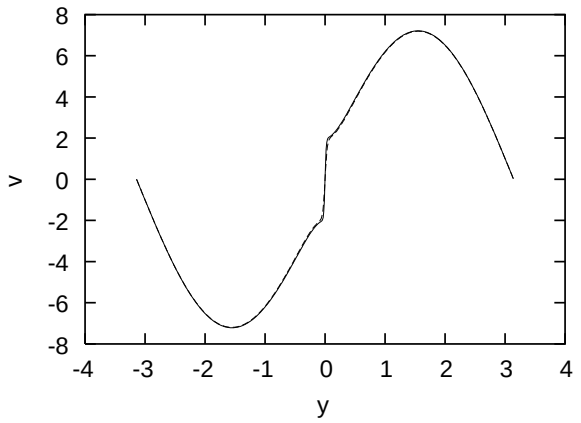


FIG. 6. Comparison of the asymptotic solution (solid line) and the numerical solution (dashed line), $\Gamma=50$, $p=25$.

$$v(x,t)\simeq -2\sqrt{\frac{v_{m1}c_0}{2\pi\beta}}\sum_{n=1}^{\infty}\left[V_n\sin\left(\frac{n\omega x}{c_0}\right)\cos(n\omega t)\right]-\frac{v_{m1}x}{L}\cos\left(\frac{\omega x}{c_0}\right)\sin(\omega t)+\frac{v_{m2}x}{L}\cos\left(\frac{2\omega x}{c_0}\right)\sin(2\omega t). \tag{75}$$

III. COMPARISON OF ASYMPTOTIC AND NUMERICAL SOLUTIONS

In this section we deal with comparison between the asymptotic (analytic) and numerical solutions of the inhomogeneous Burgers equation for stationary wave state. The accuracy of the asymptotic solutions is investigated below for various values of parameters Γ and p. The inhomogeneous Burgers equation was solved by means of the standard Runge–Kutta method of the fourth order in the frequency domain (the first 100 harmonics was used). The numerical oscillations were damped by

$$\Xi_n=\frac{\sin(nH)}{nH}, \tag{76}$$

where H is the frequency damping coefficient. Each harmonic was multiplied by the coefficient Ξ_n. It causes the additional artificial attenuation of the solution. The value of H was chosen so that the numerical oscillations practically did not arise. No damping was used for $\Gamma=20$.

The comparisons of the asymptotic solution (62) and numerical one are shown in Figs. 1–6. The first set of figures is for $\Gamma=20$ and for p equal to 0, 1 and 25, whereas the second set of figures is for $\Gamma=50$ and for the same values of the parameter p. We can observe that both solutions are in good agreement; however, the results for $\Gamma=20$ are slightly worse than for $\Gamma=50$, which is correspondence with the theoretical assumptions. In Figs. 4–6 it is difficult to distinguish the waveform of the asymptotic and numerical solution. In order to demonstrate the contribution of the new asymptotic solution (62), we compare values of harmonics for various parameters Γ in Fig. 7. For this reason we used the formula (72) with $p=0$. On the basis of this figure it is obvious that the higher harmonics differs significantly for smaller parameters Γ. Consequently, it is necessary to use, for smaller values of Γ, the asymptotic solution (62). To illustrate the validity of the spectral solution (72), we present Fig. 8. Here, we compare three waveforms. The waveform labeled by 1 represents the asymptotic solution (62) for $p=0.5$ and Γ

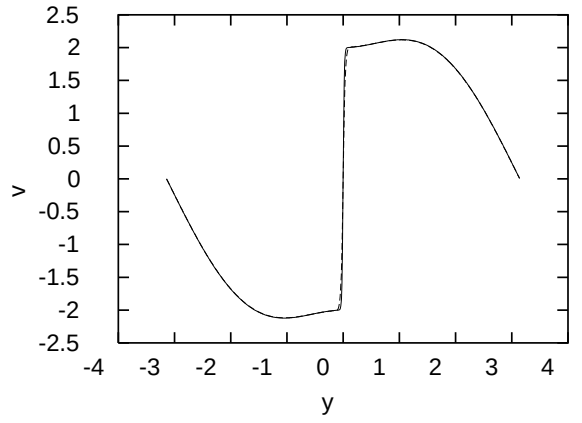


FIG. 5. Comparison of the asymptotic solution (solid line) and the numerical solution (dashed line), $\Gamma=50$, $p=1$.

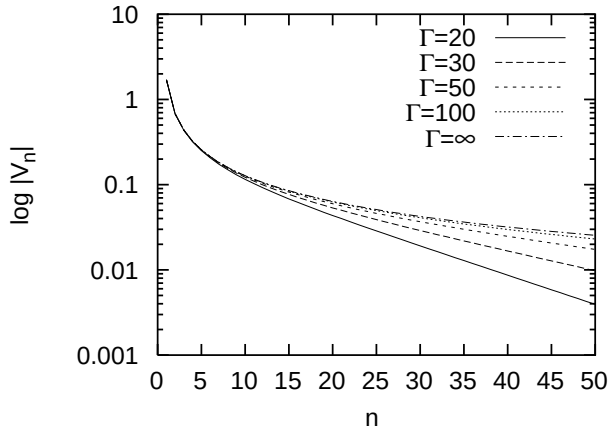


FIG. 7. Comparison of the harmonics magnitudes $|V_n|$ for various Γ and $p=0$.

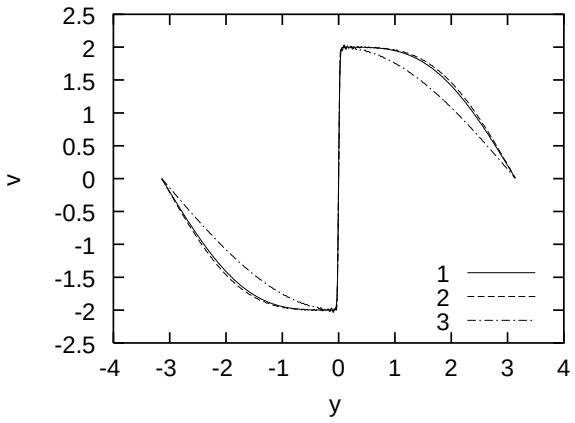


FIG. 8. Comparison of the asymptotic solution for $p=0.5$ (solid line), the spectral form solution for $p=0.5$ (dashed line), the spectral form solution for $p=0$ (dashed-dotted line), $\Gamma=50$.

$=50$, whereas the waveforms 2 and 3 are obtained from its spectral form (72) for $p=0$ and $p=0.5$. We can see that the presented limitation $p\lesssim 0.5$ enables us to get acceptable results. The waveform of (62) for $p=0$ is not depicted here because differences are not observable. The comparison between asymptotic solutions (56) and (62) is not presented here. These solutions differ significantly only for small values of Γ which do not satisfy the theoretical assumption $\Gamma \gg 1$. In spite of this fact we can say that the asymptotic solution (56) is slightly better for $\Gamma > 1$ than the solution (62), but on the other hand this solution is more complicated.

IV. CONCLUSION

The inhomogeneous Burgers equation is used for description of nonlinear standing waves in resonators of a constant diameter which are driven by a piston whose motions are characterized by two superposed sinusoidal motions. The frequency of the first motion f is equal to the resonator eigenfrequency, and the frequency of the second one is $2f$ and its phase shift is 180 deg. New asymptotic solutions of the inhomogeneous Burgers equation for stationary state regime are shown in this work. Their spectral forms are presented as well. Validity of the asymptotic solutions is verified by comparisons with numerical ones. On the basis of these comparisons it is evident that the asymptotic solutions (62) gives very good results for all $p \geqslant 0$ when Γ is sufficiently large. This solution reduces to the known solutions when $\Gamma \to \infty$ and $p \neq 0$, $\Gamma \to \infty$ and $p=0$, or $\Gamma \gg 1$ and $p=0$. The asymptotic solution (56) is slightly better than (62) for smaller values of Γ, but on the other hand this solution is more complicated. The spectral form (72) of the solution (62) enables us to obtain good results when $p \lesssim 0.5$ is satisfied. For $p=0$ the spectral form agrees with the solution (58) excellently.

The comparison between the solution for thermoviscous fluids (62) and the solutions for ideal fluids (57) illustrates that the latter fails for comparatively small values of Γ (20–50).

ACKNOWLEDGMENTS

This work was supported a grant from GACR, No. 202/01/1372, and the CTU research program J04/98:212300016.

[1] G. W. Swift, "Thermoacoustics engines," J. Acoust. Soc. Am. **84**, 1145–1180 (1988).
[2] Yu. A. Illinski, B. Lipkens, T. S. Lucas, T. W. Van Doren, and E. A. Zabolotskaya, "Nonlinear standing waves in an acoustical resonator," J. Acoust. Soc. Am. **104**, 2664–2674 (1998).
[3] W. Chester, "Nonlinear resonant oscillations of a gas in a tube of varying cross section," Proc. R. Soc. London, Ser. A **444**, 591–604 (1994).
[4] H. Ockendon, J. R. Ockendon, M. R. Peake, and W. Chester, "Geometrical effects in resonant gas oscillations," J. Fluid Mech. **257**, 201–217 (1993).
[5] M. F. Hamilton, Yu. A. Ilinskii, and E. A. Zabolotskaya, "Linear and nonlinear frequency shifts in acoustical resonators with varying cross sections," J. Acoust. Soc. Am. **110**, 109–119 (2001).
[6] V. G. Andreev, V. E. Gusev, A. A. Karabutov, O. V. Rudenko, and O. A. Sapozhnikov, "Enhancement of Q of a nonlinear acoustic resonator by means of a selectively absorbing mirror," Sov. Phys. Acoust. **31**, 162–163 (1985).
[7] O. V. Rudenko, A. L. Sobisevich, L. E. Sobisevich, and C. M. Hedberg, "Enhancement of energy and Q-factor of a nonlinear resonator with an increase in its losses," Dokl. Akad. Nauk **383**, 330–333 (2002).
[8] K. Naugolnykh and L. Ostrovsky, *Nonlinear Wave Processes in Acoustics* (Cambridge University Press, New York, 1998).
[9] N. Sugimoto, M. Masuda, T. H. Doi, and T. Doi, "Annihilation of shocks in forced oscillations of an air column in a closed tube," J. Acoust. Soc. Am. **110**, 2263–2266 (2001).
[10] P. T. Huang and J. G. Brisson, "Active control of finite amplitude acoustic waves in a confined geometry," J. Acoust. Soc. Am. **102**, 3256–3268 (1997).
[11] V. E. Gusev, H. Bailliet, P. Lotton, S. Job, and M. Bruneau, "Enhancement of the Q of a nonlinear acoustic resonator by active suppression of harmonics," J. Acoust. Soc. Am. **103**, 3717–3720 (1998).
[12] Ch. Nyberg, "Spectral analysis of a two frequency driven resonance in a closed piston," Acoust. Phys. **47**, 86–93 (1999).
[13] W. Chester, "Resonant oscillations in closed tubes," J. Fluid Mech. **18**, 44–64 (1964).
[14] V. Kaner, O. V. Rudenko, and R. Khokhlov, "Theory of nonlinear oscillations in acoustic resonators," Sov. Phys. Acoust. **23**, 432–437 (1977).
[15] V. E. Gusev, "Buildup of forced oscillations in acoustic resonator," Sov. Phys. Acoust. **30**, 121–125 (1984).
[16] M. Ochmann, "Nonlinear resonant oscillations in closed tubes–An application of the averaging method," J. Acoust. Soc. Am. **77**, 61–66 (1985).
[17] O. V. Rudenko, C. M. Hedberg, and B. O. Enflo, "Nonlinear standing waves in a layer excited by the periodic motion of its boundary," Acoust. Phys. **47**, 452–460 (2001).
[18] A. B. Coppens and J. V. Sanders, "Finite-amplitude standing waves within real cavities," J. Acoust. Soc. Am. **58**, 1133–1140 (1975).
[19] D. B. Cruishank, "Experimental investigation of finite-amplitude acoustic oscillations in a closed tube," J. Acoust. Soc. Am. **52**, 1024–1036 (1972).
[20] C. Lawrenson, B. Lipkens, T. S. Lucas, D. K. Perkins, and T. W. Van Doren, "Measurements of macrosonic standing waves in oscillating closed cavities," J. Acoust. Soc. Am. **104**, 623–636 (1998).
[21] V. P. Kuznetsov, "Equations of nonlinear acoustics," Sov. Phys. Acoust. **16**, 467–470 (1971).
[22] M. F. Hamilton and D. T. Blackstock, *Nonlinear Acoustic* (Academic, New York, 1998).
[23] K. M. Urwin and F. M. Arscott, "Theory of the Whittaker–Hill equation," Proc. R. Soc. Edinburgh, Sect. A: Math. Phys. Sci. **69**, 28–44 (1970).
[24] K. M. Urwin, "Characteristic values of the Whittaker–Hill equation with one large parameter," Bul. Institut Politehnic Iasi **26**, 33–40 (1975).
[25] N. W. McLachlan, *Theory and Application of Mathieu Functions* (Dover, New York, 1964).
[26] V. E. Gusev, "Parametric amplification of acoustic waves in closed tubes," Sov. Phys. Acoust. **30**, 176–180 (1984).
[27] A. A. Karabutov, E. A. Lapshin, and O. V. Rudenko, "Interaction between light waves and sound under acoustic nonlinearity conditions," Sov. Phys. JETP **44**, 58–63 (1976).

Steady state unfocused circular aperture beam patterns in nonattenuating and attenuating fluids

Albert Goldstein[a)]
Department of Radiology, Wayne State University School of Medicine, Detroit Receiving Hospital, Detroit, Michigan 48201

(Received 7 June 2003; revised 24 September 2003; accepted 13 October 2003)

Single integral approximate formulas have been derived for the axial and lateral pressure magnitudes in the beam pattern of steady state unfocused circular flat piston sources radiating into nonattenuating and attenuating fluids. The nonattenuating formulas are shown to be highly accurate at shallow beam depths if a normalized form of the beam pattern is utilized. The axial depth of the beginning of the nonattenuated beam pattern far field is found to be at $6.41Y_0$. It is demonstrated that the nonattenuated lateral beam profile is represented at this and deeper depths by a Jinc function directivity term. Values of α and z are found that permit the attenuated axial pressure to be represented by a plane wave multiplicative attenuation factor. This knowledge should aid in the experimental design of high accuracy attenuation measurements. The shifts in depth of the principal axial pressure maxima and minima due to fluid attenuation are derived. Single integral approximate equations for the attenuated full beam pattern pressure are presented using complex Bessel functions. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1631286]

PACS numbers: 43.20.Rz, 43.20.El, 43.30.Jx [SFW] Pages: 99–110

I. INTRODUCTION

Ultrasound waves generated by unfocused circular plane piston sources are routinely used in acoustic measurements and have been extensively studied.[1,2] Diffraction effects in attenuation measurements using these sources also have been investigated.[3,4]

To aid in the use of these important acoustic sources in fluids, new easy to use approximate single integral expressions have been derived. Human tissue, like fluids, only supports longitudinal wave propagation[5] and will be used in the presented numerical examples.

Axial and full beam patterns in nonattenuating fluids will be studied first. Then axial and full beam patterns in attenuating fluids will be investigated. Since most attenuation measurements are performed along the unfocused beam axis, knowledge of the axial beam pattern is essential in interpreting experimental results.

II. NONATTENUATING FLUIDS

The circular aperture piston source is surrounded by an infinite rigid baffle. It is driven by steady state or tone burst driving voltages. The radiating surface moves uniformly with speed $u_0 \exp(i\omega t)$ normal to its surface. The fluid is homogeneous, istotropic, nonscattering, and nonattenuating. The pressure amplitude is assumed low enough to avoid cavitation and/or nonlinear effects in the fluid. And the piston source is assumed to have been designed[6] to avoid Lamb wave propagation on its radiating surface.[7]

Figure 1 presents the beam geometry. The circular plane piston source is in the XY plane centered at the origin. Without loss of generality, the field point Q is in the XZ plane a lateral distance x from the Z axis (beam axis) at an axial depth z. Each point dS on the source surface emits spherical waves. Integration over all of these point sources, each with strength $dq = u_0\, dS = u_0\, dx\, dy = u_0 \sigma\, d\sigma\, d\varphi$, produces a pressure amplitude $P(r,\theta,t)$ at Q a distance r from the origin at an angle θ from the Z axis. The pressure at point Q is[8]

$$P(r,\theta,t) = i\frac{\rho_0 \omega u_0}{2\pi}\int_0^{2\pi}\int_0^a \frac{e^{i(\omega t - kr')}}{r'}\sigma\, d\sigma\, d\varphi, \tag{1}$$

where $k = 2\pi/\lambda = \omega/c$, ω is the angular frequency, λ is the wavelength, a is the piston radius, c is the fluid acoustic velocity, and ρ_0 is the mean physical mass density of the fluid. Due to the source circular symmetry, the resulting beam pattern has cylindrical symmetry.

A. Axial pressure

Along the beam axis $r = z$ and $\theta = 0$. Equation (1) reduces to

$$P(z,0,t) = i\rho_0 \omega u_0 e^{i\omega t}\int_0^a \frac{e^{-ik\sqrt{z^2+\sigma^2}}}{\sqrt{z^2+\sigma^2}}\sigma\, d\sigma \tag{2}$$

which, due to the high beam symmetry, is an exact integral yielding a complex acoustic pressure

$$P(z,0,t) = \rho_0 c u_0 e^{i\omega t}[e^{-ikz} - e^{-ik\sqrt{z^2+a^2}}], \tag{3}$$

that can be rewritten in the form

$$P(z,0,t) = \rho_0 c u_0 e^{i(\omega t - kz)}[1 - e^{-ik(\sqrt{z^2+a^2}-z)}]. \tag{4}$$

This last expression contains three multiplicative terms. The middle exponential term represents a one-dimensional traveling unity amplitude plane wave. The first and third terms represent the diffraction caused by the unfocused plane piston beam pattern. At any time t the pressure magnitude in the unfocused beam pattern may be determined by calculat-

a)Electronic mail: agoldste@med.wayne.edu

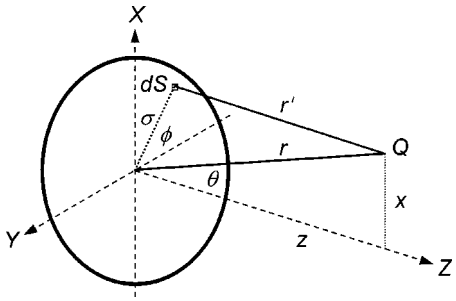


FIG. 1. Beam geometry of a radiating circular flat piston of radius a. The piston face is in the XY plane. Q is the field point. The source point $u_0\,dS$ is on the flat piston surface σ from its center at an angle φ from the Y axis.

ing the magnitude of the product of the first and third terms in Eq. (4),

$$|P(z,0)| = 2\rho_0 c u_0 \left| \sin\left\{ \frac{k}{2}\left(\sqrt{z^2+a^2}-z\right)\right\}\right|. \tag{5}$$

In the ultrasound literature the Fraunhofer zone beyond the last axial maximum at $Y_0 = a^2/\lambda$ is called the ''far zone.'' This is not to be confused with the ''far field,'' which is that range of distances from the source where it is effectively a point source producing spherical waves with a $1/r$ amplitude dependence.

To calculate the far zone axial depth of the beginning of the far field of an unfocused plane piston source start with Eq. (5). At large z the argument of the sine term goes to zero and $\sin(x)$ may be replaced by x. Expanding the square root in the argument of the sine using a binomial series (which is valid when $z > a$),

$$\sin\left\{ \frac{k}{2}\left(\sqrt{z^2+a^2}-z\right)\right\} \approx \sin\left\{ \frac{ka^2}{4z}\right\} \approx \frac{\pi a^2}{2\lambda z} \tag{6}$$

producing the required spherical waves. The shortest axial depth where the right-hand side of Eq. (6) is valid may be found using the first two terms of the series expansion of $\sin(x)$,

$$\sin(x) = x - \frac{x^3}{6} + \cdots \tag{7}$$

and normalizing z with respect to the distance Y_0

$$z = n\frac{a^2}{\lambda}, \tag{8}$$

where n is a positive number. $\mathrm{Sin}(x) = x$ is true to 1% when the second term in Eq. (7) is 1% of the first term. In that case $n = 6.41$. So the far field exists at axial depths greater than 6.41 times Y_0. The practical relevance of determining the depth of the beginning of the far field will be seen in the discussion below of the Jinc Function approximation of the lateral beam profile.

B. Full beam pattern

Calculating the pressure lateral to the beam axis at all depths requires evaluating Eq. (1) at each field point Q. r' is replaced with the geometric identity[9]

$$r' = (r^2 + \sigma^2 - 2r\sigma\sin\theta\cos\varphi)^{1/2} \tag{9}$$

to obtain the exact full beam pattern pressure. The Fresnel approximation used to evaluate Eq. (1) involves replacing r' in the denominator of the integral's kernel with r and taking it out of the integral. Then the square root in Eq. (9) is expanded using a binomial series keeping only the first two terms

$$r' \cong r\left(1 + \frac{\sigma^2}{2r^2} - \frac{\sigma\sin\theta\cos\varphi}{r}\right) \tag{10}$$

and this expression is used in the exponent of the integral kernel. The next terms in the binomial series that were neglected in obtaining Eq. (10) are about three orders of magnitude smaller than the two terms that were kept.

A standard approximation is to neglect the second term in the parentheses in Eq. (10) at considerable distances from the source.[8] From Fig. 1 $\sin\theta = x/r$. For field points within a distance a of the beam axis, the ratio σ/r is roughly equal to $\sin\theta$. And for field points further than a from the beam axis, the ratio σ/r is less than $\sin\theta$. So the second and third terms inside the parentheses in Eq. (10) have approximately the same magnitude regardless of distance from the source.

Keeping all three terms the approximate pressure is

$$P_{\mathrm{ap}}(r,\theta,t) = i\frac{\rho_0\omega u_0}{2\pi r}e^{i(\omega t - kr)}\int_0^a \sigma e^{-i(k\sigma^2/2r)} \times \int_0^{2\pi} e^{ik\sigma\sin\theta\cos\varphi}\,d\varphi\,d\sigma. \tag{11}$$

Using the identity[9]

$$2\pi J_m(x) = (-i)^m \int_0^{2\pi} e^{ix\cos\varphi}\cos(m\varphi)\,d\varphi, \tag{12}$$

where $m = 0,1,2,3...$, Eq. (11) becomes

$$P_{\mathrm{ap}}(r,\theta,t) = i\frac{\rho_0\omega u_0}{r}e^{i(\omega t - kr)}\int_0^a \sigma e^{-i(k\sigma^2/2r)} \times J_0(k\sigma\sin\theta)\,d\sigma. \tag{13}$$

Here a unity amplitude traveling spherical wave is modulated by two beam pattern diffraction terms. At any time t the magnitude of the unfocused beam pattern may be determined by calculating the magnitude of the product of the beam pattern diffraction terms in Eq. (13),

$$|P_{\mathrm{ap}}(r,\theta)| = \frac{\rho_0\omega u_0}{r}\{I_{\mathrm{C}}^2(r,\theta) + I_{\mathrm{S}}^2(r,\theta)\}^{1/2}, \tag{14}$$

where

$$I_{\mathrm{C}}(r,\theta) = \int_0^a \sigma\cos\left(\frac{k\sigma^2}{2r}\right)J_0(k\sigma\sin\theta)\,d\sigma \tag{15}$$

and

$$I_{\mathrm{S}}(r,\theta) = \int_0^a \sigma\sin\left(\frac{k\sigma^2}{2r}\right)J_0(k\sigma\sin\theta)\,d\sigma. \tag{16}$$

Equations (11), (13), and (14) can be used to compute the beam pattern of an unfocused piston source in a nonattenuating fluid. Their accuracy can be verified by comparison to both the exact pressure along the beam axis [Eq. (5)] and

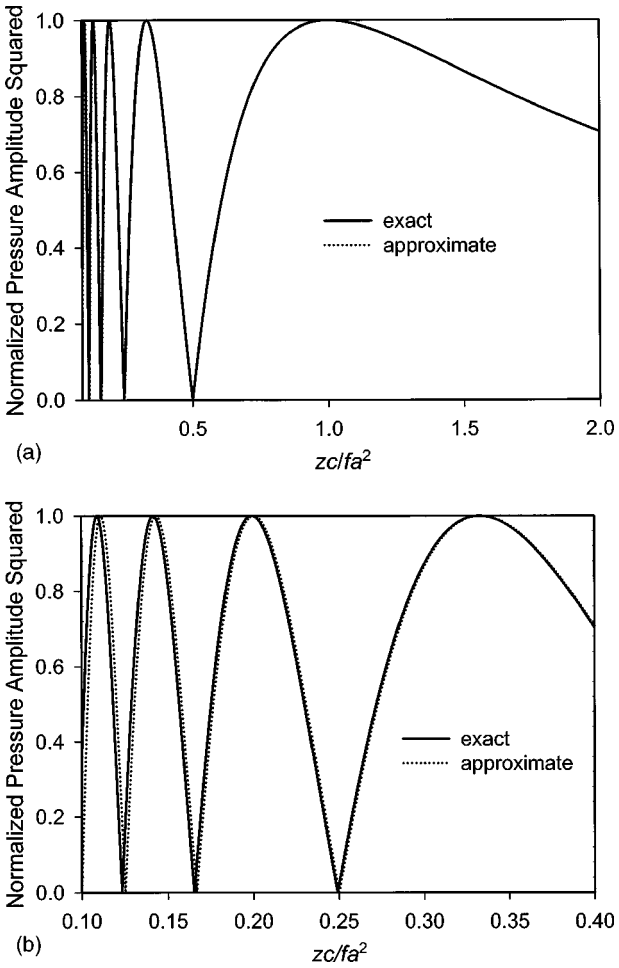


FIG. 2. Check of the accuracy of the approximate expression for the beam pattern pressure. The square of the approximate beam pattern axial pressure [Eq. (19)] is plotted along with the square of the exact expression for the axial pressure [Eq. (5)] for $a=1$ cm, $f=5$ MHz, and $c=0.154$ cm/μs. Both pressures are normalized to unity. (a) At the end of the near zone and in the far zone the approximate expression agrees with the exact expression. (b) In the near zone at depths less than $0.2Y_0$ the approximate expression begins to differ from the exact expression.

the exact equations in Appendix D (with $\alpha=0$).

Along the beam axis $r=z$ and $\theta=0$. The two integrals in Eq. (11) are easily integrated to obtain

$$P_{\mathrm{ap}}(z,0,t)=-\rho_0 c u_0 e^{i(\omega t-kz)}[e^{-(ika^2/2z)}-1]. \tag{17}$$

The pressure magnitude along the beam axis is then

$$|P_{\mathrm{ap}}(z,0)|=\rho_0 c u_0\left\{2\left(1-\cos\left(\frac{ka^2}{2z}\right)\right)\right\}^{1/2}. \tag{18}$$

Equation (18) also can be expressed in the form

$$|P_{\mathrm{ap}}(z,0)|=2\rho_0 c u_0\left|\sin\left(\frac{ka^2}{4z}\right)\right|. \tag{19}$$

Equation (19) is seen to be an approximate form of Eq. (5) when the binomial series approximation of Eq. (6) is valid.

The exact [Eq. (5)] and approximate [Eq. (19)] expressions for the magnitude of the axial pressure may be compared as an indication of the accuracy of the approximate full beam pattern pressure [Eq. (14)] in the near zone. This comparison is performed in Figs. 2(a) and (b) for $a=1$ cm, $f=5$ MHz and $c=0.154$ cm/μs. Here the distance axis is represented by the normalized variable zc/fa^2 (i.e., in units of Y_0). For these values of a, f, and c the approximate expression is a very good fit to the exact expression except at axial distances closer than $0.2Y_0$.

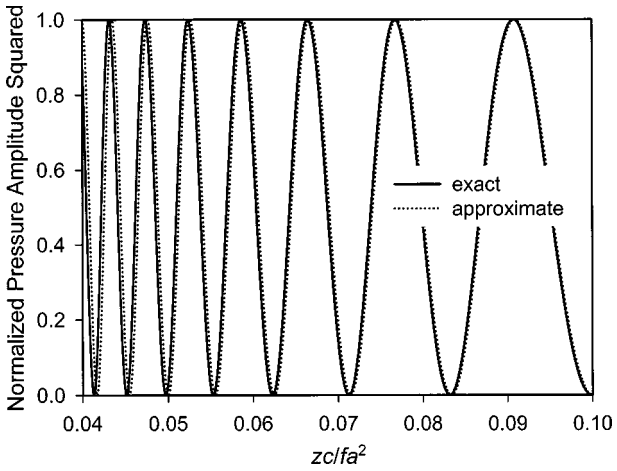


FIG. 3. Check of the accuracy of the approximate expression for the beam pattern pressure. The square of the approximate beam pattern axial pressure [Eq. (19)] is plotted along with the square of the exact expression for the axial pressure [Eq. (5)] for $a=4$ cm, $f=5$ MHz, and $c=0.154$ cm/μs. Both pressures are normalized to unity. Compared to Fig. 2(b) the approximate axial expression agrees with the exact expression as close as $0.05Y_0$ as predicted by Eq. (21).

The exact expressions in Appendix D (with $\alpha=0$) could be used to compute the beam pressure closer than $0.2Y_0$. However, the approximate Eq. (14) can be used to higher accuracy by taking advantage of the normalized representation of the unfocused, unattenuated beam pattern[1] with the variable zc/fa^2 in the axial direction and x/a in the lateral direction.

In Fig. 2(b) for illustration purposes the axial distance $0.2Y_0$ was arbitrarily selected as a "good fit" between the approximate and exact axial pressure expressions. Due to the binomial expansion used [cf. Eq. (6)], this "good fit" is uniquely related to the ratio ξ between the axial distance z and the source radius a

$$z=\xi a. \tag{20}$$

In Fig. 2(b) $Y_0=32.468$ cm and $z=0.2Y_0=6.494$ cm. So $\xi=6.494$ for a "good fit." Substituting Eq. (20) into Eq. (8) and solving for n,

$$n\geqslant\frac{\xi c}{af}. \tag{21}$$

So the "good fit," related to this determined ξ value, can be attained at any value of n by suitable selection of c, a, and f.

Figure 3 is a comparison of the approximate and exact axial pressure equations for $a=4$ cm, $f=5$ MHz, and $c=0.154$ cm/μs. The n value for a "good fit" has reduced to 0.05 as predicted by Eq. (21).

Thus, the normalized representation of an unfocused steady state circular aperture beam pattern in the near zone can be computed using Eq. (14) to arbitrary accuracy by properly selecting values of a, f, and c. Once this accurate

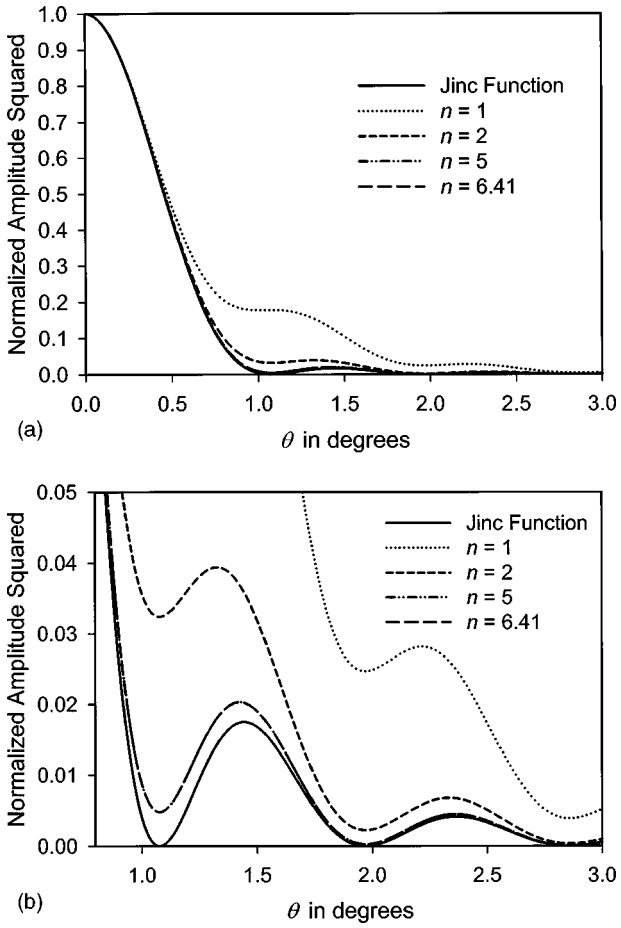


FIG. 4. Comparison of the lateral beam profile to the Jinc function directivity term in a pulse-echo tone burst measurement at various depths, nY_0, in the beam pattern. The squares of the pressure and Jinc function are plotted vs the angle θ in degrees. (a) At the depth of Y_0 ($n=1$) the first axial minimum does not exist. (b) A magnified portion of (a) demonstrates clearly that at the depth $6.41Y_0$, which is the onset of the far field beam pattern, the two are identical. For $n>1.15$ the angular position of the first minimum in the lateral beam profile is identical to the first lateral zero of the Jinc function directivity term. See text for details.

normalized beam pattern is obtained it easily can be scaled back to real-space dimensions using the a and f of the source and the acoustic velocity of the fluid.

Acoustics texts represent the unfocused circular piston far zone beam pattern by a Jinc function directivity term

$$\frac{2J_1(ka\sin\theta)}{ka\sin\theta}. \tag{22}$$

The appropriateness of using this Jinc function expression for the lateral beam profile in the far zone may now be examined. Equation (14) was used to compute the lateral beam profile for $a=1$ cm, $f=5$ MHz, and $c=0.154\,\mathrm{cm}/\mu\mathrm{s}$ at beam pattern depths of nY_0 with $n=1$, 2, 5, and 6.41. The results are shown in Figs. 4(a) and (b) where the square of the Jinc function is compared to the square of the pressure. This was done to avoid the oscillations of the Jinc function around zero pressure [Eq. (14) does not oscillate since it represents the magnitude of the pressure] and to simulate a pulse-echo tone burst lateral beam profile measurement.

The lateral beam profile at each depth was normalized to its central, axial amplitude and, for convenience, the independent variable is the angle θ. Figure 4(a) demonstrates that as n increases from unity the lateral beam profile rapidly approaches a Jinc function. Figures 4(a) and (b) demonstrate that the $n=6.41$ lateral beam profile is identical to the Jinc function directivity term. So the Jinc function directivity term represents the lateral beam profile in the far field. Note that Figs. 4(a) and (b) present the lateral beam profile at constant z. If the variable r in Eqs. (14), (15), and (16) had been kept constant while the angle θ varied, the far zone directivity function would have been plotted.

Since the lateral beam profile is effectively a Jinc function in the far field, the second term in the parentheses in Eq. (10) is negligible at these depths. Appendix A demonstrates this in detail. This term cannot be neglected due to its relative magnitude compared to the third term in the parentheses in Eq. (10) as discussed above. It can be neglected at large depths in the far zone ($z\geqslant 6.41Y_0$) because of the absolute magnitude of its resultant trigonometric function.

Experimenters sometimes use the first lateral zero of an assumed far zone Jinc function directivity term to determine the effective radius of a plane piston source. The question arises if this procedure is accurate for lateral beam profile measurements at depths less than $6.41Y_0$. Figures 4(a) and (b) demonstrate that for $n<6.41$ the lateral beam profile has a minimum close to zero magnitude. For $n=1$ this minimum does not exist.

The locations of the lateral beam profile minimum for various n values were found numerically by plotting the first derivative of the pressure squared [Eq. (14)] (Mathcad, Mathsoft, Cambridge, MA) and noting where the first derivative crossed the zero axis. This derivative exhibited a minimum for $n=1.11$ and greater. For $n>1.15$ the pressure squared minimum was found to have exactly the same x (or x/a) value as the Jinc function directivity term first lateral zero. So the assumption of a far zone Jinc function lateral beam profile (or directivity function) would yield an accurate value for the plane piston source effective radius when $n>1.15$.

The effect on the beam pattern of radial apodization or amplitude shading on the source face may be calculated by inserting a source amplitude radial shading factor into the kernels of the integrals in Eqs. (11) and (D1). Or the beam pattern of an unfocused biopsy transducer with a central hole may be calculated by changing the lower radius limit in these integrals.

III. ATTENUATING FLUIDS

Equation (1) may be generalized to propagation in an attenuating fluid by replacing k with the complex wave vector

$$k_c=k-i\alpha, \tag{23}$$

whose imaginary component is the fluid's attenuation coefficient in units of Nepers/cm. The substitution of k_c for k must be made very early in the mathematical development of the pressure otherwise there is a possibility that this substitution may not be made properly. An improper substitution would represent a change in the attenuation properties of the fluid

during the course of the calculation of the attenuated pressure.

The best place to make the substitution of k is in the most basic equation for the pressure, e.g., Eq. (1). It is possible to make this substitution properly in subsequent equations in the mathematical development as long as k remains a constant. Any mathematical operation involving frequency would not preserve k because α may have a different frequency dependence than k. This late substitution procedure is valid for a constant frequency CW source. There is a simple test to determine if the substitution may be made in a later step in a calculation. Starting from an expression known to yield the correct result when Eq. (23) is substituted, perform the next computation step twice—computation then substitute or substitute then computation. If the results of the two methods are identical then the substitution of Eq. (23) can be properly made from the later theoretical expression.

A. Axial pressure

For plane or spherical wave propagation fluid attenuation results in a multiplicative exponential attenuation factor (see Appendix B). Attenuation measurements performed along the beam axis of a circular aperture piston source are interpreted by assuming the existence of this factor. It is important to determine under what conditions this simple multiplicative exponential attenuation factor is an accurate representation of the attenuated pressure.

While the substitution indicated in Eq. (23) should be made in Eq. (1), it may be made in Eq. (3) if the factor k from the integration had been preserved

$$P_\alpha(z,0,t)=\frac{\rho_0\omega u_0}{k-i\alpha}e^{i\omega t}[e^{-ikz}e^{-\alpha z} - e^{-ik\sqrt{z^2+a^2}}e^{-\alpha\sqrt{z^2+a^2}}]. \tag{24}$$

This can be set in the form

$$P_\alpha(z,0,t)=\frac{\rho_0\omega u_0(k+i\alpha)}{k^2+\alpha^2}e^{i(\omega t-kz)} \times[e^{-\alpha z}-e^{-ik(\sqrt{z^2+a^2}-z)}e^{-\alpha\sqrt{z^2+a^2}}], \tag{25}$$

where a propagating unity amplitude one-dimensional traveling plane wave term is modulated by beam pattern diffraction terms.

An exact expression for the attenuated axial pressure magnitude is obtained from the magnitude of the product of the beam pattern diffraction terms in Eq. (25),

$$|P_{\alpha\mathrm{EX}}(z,0)|=\frac{\rho_0\omega u_0}{\sqrt{k^2+\alpha^2}}e^{-\alpha z}\{1+e^{-2\alpha(\sqrt{z^2+a^2}-z)} - 2e^{-\alpha(\sqrt{z^2+a^2}-z)} \times\cos(k(\sqrt{z^2+a^2}-z))\}^{1/2}. \tag{26}$$

For low attenuation and/or large axial distances the two exponential terms inside the curly brackets go to unity producing an approximate expression for the axial pressure magnitude in an attenuating fluid

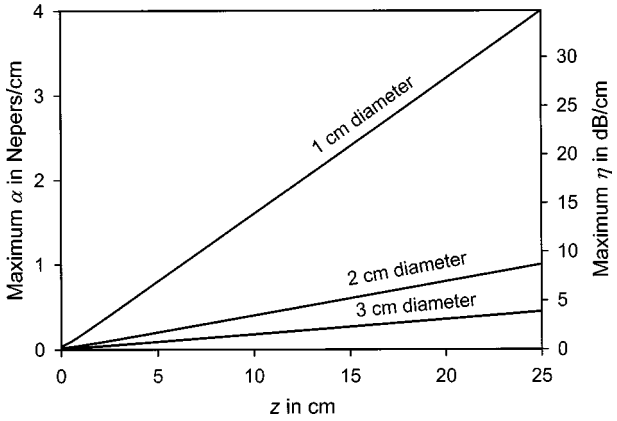


FIG. 5. Maximum attenuation for 1% accuracy of Eq. (27) for $P_{\alpha\mathrm{APP}}$ (i.e., $\kappa=4$). The exact portion of Eq. (28) was used to produce this plot. At each axial depth z any attenuation value below the line for each piston diameter will ensure that $P_{\alpha\mathrm{APP}}$ is accurate to better than 1% compared to $P_{\alpha\mathrm{EX}}$. The lower the attenuation the more accurate is $P_{\alpha\mathrm{APP}}$.

$$|P_{\alpha\mathrm{APP}}(z,0)|=\frac{2\rho_0\omega u_0}{\sqrt{k^2+\alpha^2}}e^{-\alpha z}\sqrt{\sin^2\left\{\frac{k}{2}(\sqrt{z^2+a^2}-z)\right\}}. \tag{27}$$

When the approximate Eq. (27) is valid the fluid attenuation appears in the usually assumed form of a multiplicative exponential attenuation factor. If Eq. (8) is substituted into the exponential terms in either Eq. (26) or Eq. (27) it can easily be shown that the normalized representation[1] valid for an unattenuated beam pattern is not valid when the fluid is attenuating.

By considering the exponential terms in the exact Eq. (26) it can be determined what values of α and z permit use of the approximate Eq. (27). The first two terms in the binomial series expansion of e^x are $1+x$, so for κ% accuracy in assuming that an exponential term goes to unit x must be less than $\kappa/100$. The second term in the curly brackets in Eq. (26) has the larger exponential argument, so for κ% accuracy in replacing this exponential term (and $\kappa/2$% accuracy in replacing the other exponential term) by unity the condition on α and z is

$$\alpha\leqslant\frac{\kappa}{200(\sqrt{z^2+a^2}-z)}\approx\frac{\kappa z}{100a^2} \tag{28}$$

with the binomial series approximation on the right-hand side valid when $z>a$.

When the exponential terms in Eq. (26) are replaced with $1-\kappa/100$ and $1-\kappa/200$, respectively, it is found that

$$|P_{\alpha\mathrm{EX}}(z,0)|=\sqrt{1-\frac{\kappa}{200}}|P_{\alpha\mathrm{APP}}(z,0)|. \tag{29}$$

For small κ values the square root in Eq. (29) may be replaced with the first two terms of its binomial series expansion and the percentage error in replacing Eq. (26) by Eq. (27) is then $\kappa/4$%.

When there is an equals sign on the left of Eq. (28) α has its largest value consistent with Eq. (27) being within $\kappa/4$% of Eq. (26). Figure 5 presents this maximum α value as a function of z for three different circular aperture piston

diameters. It demonstrates that as z increases, larger maximum values of α are permitted without causing more than a 1% error. Since α is proportional to κ in Eq. (28), to represent the exact axial pressure to 10% accuracy both ordinates in Fig. 5 should be multiplied by 10.

To determine the validity of Eq. (27) in the region of Y_0, note that at the last axial maximum the argument of the sine term in Eq. (27) is $\pi/2$. Then, at the last axial maximum

$$\sqrt{z^2+a^2}-z=\frac{\lambda}{2} \tag{30}$$

so the argument of the first exponential term in Eq. (26) is $-\alpha\lambda$. If $\alpha\lambda$ is small, Eq. (27) will be accurate to $\alpha\lambda/4\%$ at Y_0. This reduces to $\alpha_0 c/4\%$ in tissue where $\alpha \approx f\alpha_0$ (see Appendix B).

The fluid attenuation may be represented by a multiplicative exponential attenuation factor when the product of attenuation and acoustic path length from all point sources on the flat piston face to the field point on the beam axis are approximately equal. The larger the piston diameter, the more the point source acoustic path lengths will vary. So, the maximum attenuation value for $P_{\alpha\mathrm{APP}}(z,0)$ to be within 1% of $P_{\alpha\mathrm{EX}}(z,0)$ will be lower with larger piston diameter as is seen in Fig. 5. Also, the acoustic path lengths will be closer in magnitude the further away the field point. Figure 5 demonstrates this because the maximum attenuation for the multiplicative exponential attenuation factor to be valid increases with depth. Conventional wisdom has been that with high attenuation beam diffraction effects are minimal. Figure 5 demonstrates that this is not the case, especially at short distances.

It is instructive to study the differences between $|P_{\alpha\mathrm{EX}}(z,0)|$ and $|P_{\alpha\mathrm{APP}}(z,0)|$ at various levels of fluid attenuation. Figures 6(a), (b), and (c) demonstrate the differences in axial pressure magnitude between these exact and approximate expressions for a 4 MHz unfocused circular aperture plane piston source with a 1 cm radius transmitting into a fluid with $c=0.154\,\mathrm{cm}/\mu\mathrm{s}$. A low attenuation of 0.5 dB/(cm MHz) [Fig. 6(a)] results in higher order extrema (those closer to the source face) with progressively more of a difference between the exact and approximate expressions. Also, the approximate expression always has zero valued minima, while the exact expression minima closer to the source face differ from zero.

With a much higher attenuation of 5.2 dB/(cm MHz) these differences increase with the exact expression now resembling an exponentially decaying pressure amplitude with superimposed oscillations as demonstrated in the semilog plot of Fig. 6(b). The approximate maxima lie on a straight line in this plot and the exact maxima lie on a slightly curved line. Since the exact expression represents the data which would be measured in this situation, an experimenter would make an error by interpreting the near zone data as being caused by a multiplicative attenuation factor as in Appendix B [e.g., Eq. (27)]. This is another demonstration of the error in assuming that with high attenuation beam diffraction may be neglected.

Figure 6(c) shows the two pressure equations in Fig. 6(b) at the last axial maximum. At a slightly higher fluid attenuation the last axial maximum and minimum will disappear for the exact equation. Since this occurs at a pressure amplitude approximately 292 dB down for transmission it has little practical significance.

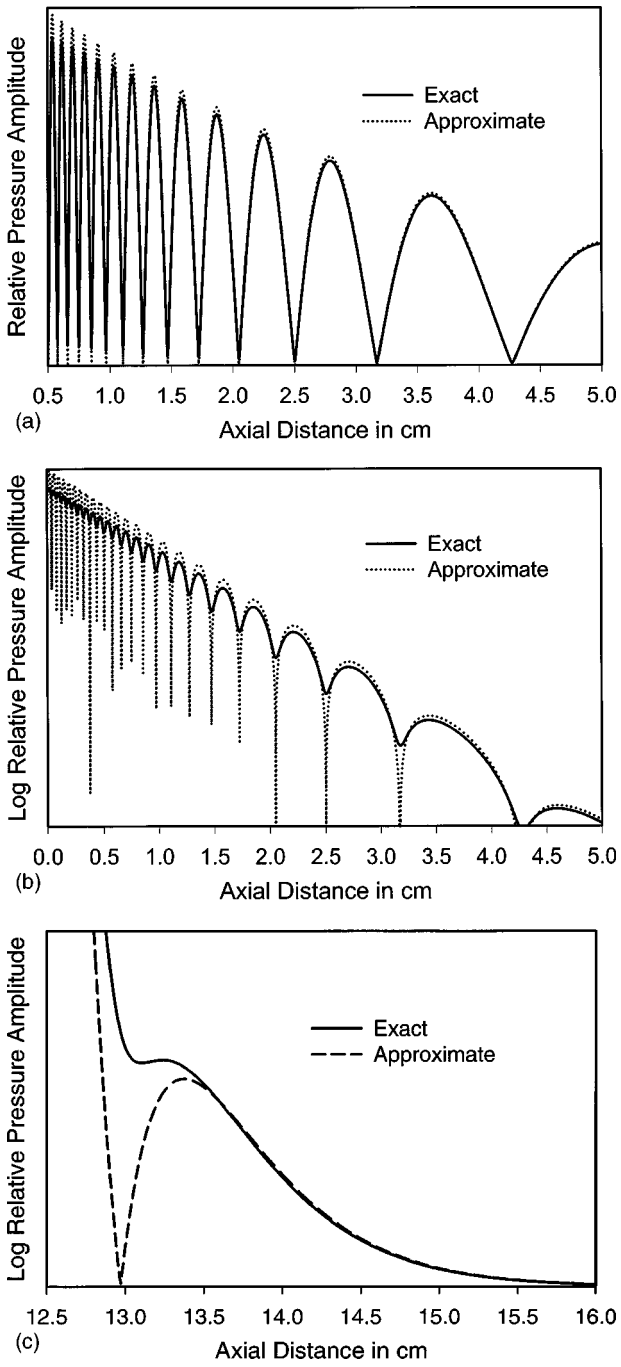


FIG. 6. Comparison of the exact and approximate expressions for the attenuated axial pressure. Equations (26) and (27) are plotted for $f=4$ MHz, $a=1$ cm, and $c=0.154\,\mathrm{cm}/\mu\mathrm{s}$. (a) Relative linear plot for $\eta_0=0.5$ dB/(cm MHz) and (b) relative log plot for $\eta_0=5.2$ dB/(cm MHz). The approximate expression becomes less accurate closer to the source. The minima of the approximate expression should always go to zero, but sampling difficulties prevented them from reaching zero in this numerical plot. (c) Relative log plot of the last axial maximum for $\eta_0=5.2$ dB/(cm MHz). At larger values of η_0 the exact expression no longer has this peak.

An important effect attenuation has on the axial pressure maxima is a distortion of their shape leading to shifts of their axial positions. To calculate these shifts in axial maxima po-

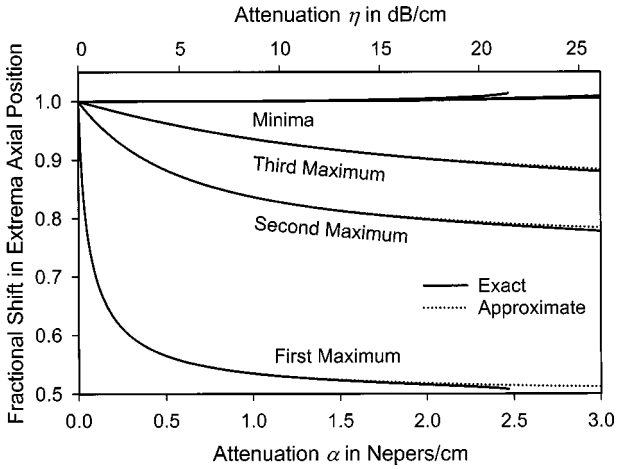


FIG. 7. Shift of the axial extrema due to attenuation. The maxima are shifted closer to the source due to their broad peaks. The positions of the sharp minima are mostly unaffected by the attenuation. The first maximum (or Y_0 peak) is the most affected by attenuation. The Y_0 peak disappears in the exact expression at high attenuation as demonstrated in Fig. 6(c). At these higher attenuations the approximate expression is invalid for the Y_0 peak.

sition Eqs. (26) and (27) are differentiated and set equal to zero. Differentiating Eq. (27) for $|P_{\alpha\mathrm{APP}}(z,0)|$ and setting the result equal to zero

$$\tan\left\{\frac{k}{2}(\sqrt{z^2+a^2}-z)\right\}=\frac{k}{2\alpha}\left(\frac{z}{\sqrt{z^2+a^2}}-1\right). \tag{31}$$

And differentiating Eq. (26) for $|P_{\alpha\mathrm{EX}}(z,0)|$ and setting the result equal to zero

$$\begin{aligned}&\alpha e^{\alpha(\sqrt{z^2+a^2}-z)}+\frac{\alpha z}{\sqrt{z^2+a^2}}e^{-\alpha(\sqrt{z^2+a^2}-z)}\\&=\alpha\left(\frac{z}{\sqrt{z^2+a^2}}+1\right)\cos\{k(\sqrt{z^2+a^2}-z)\}\\&\quad+k\left(\frac{z}{\sqrt{z^2+a^2}}-1\right)\sin\{k(\sqrt{z^2+a^2}-z)\}.\end{aligned} \tag{32}$$

These two equations were solved by numerical methods (Mathcad, Mathsoft, Cambridge, MA). The results of these computations for the first three axial maxima and minima are presented in Fig. 7 for the same case as Fig. 6 ($a=1$ cm, $f=4$ MHz, and $c=0.154\,\mathrm{cm}/\mu\mathrm{s}$). The sharp axial minima are not shifted. But the broad axial maxima are shifted to shorter axial distances by the attenuation of the fluid. In Fig. 7 the first maximum is the Y_0 maximum. Since it is the broadest, it is the most affected. It has been demonstrated previously that even the low attenuation of water can cause a 1% shift in position of Y_0.[6]

In Fig. 7 the $P_{\alpha\mathrm{EX}}$ and $P_{\alpha\mathrm{APP}}$ plots yield essentially identical results for all but the highest values of attenuation. The first maximum has different results around 5 dB/(cm MHz)$\times$4 MHz$=$20 dB/cm as seen in Fig. 6(c). The first maximum $P_{\alpha\mathrm{EX}}$ peak disappears at 2.47 Np/cm or 21.45 dB/cm so the first maximum $P_{\alpha\mathrm{APP}}$ curve at higher attenuations is not valid.

B. Full beam pattern

The attenuated full beam pattern is derived from Eq. (1) with the substitution of Eq. (23) and then applying a Fresnel approximation for r'. Or the Fresnel approximation may be applied first and then the substitution indicated in Eq. (23) made without error. So Eq. (23) may be substituted into the result of the Fresnel approximation [Eq. (11)] to obtain

$$\begin{aligned}&P_{\alpha\mathrm{APP}}(r,\theta,t)\\&=i\frac{\rho_0\omega u_0}{2\pi r}e^{i(\omega t-kr)}e^{-\alpha r}\int_0^a\sigma e^{-i(k\sigma^2/2r)}e^{-(\alpha\sigma^2/2r)}\\&\quad\times\int_0^{2\pi}e^{ik\sigma\sin\theta\cos\varphi}e^{\alpha\sigma\sin\theta\cos\varphi}\,d\varphi\,d\sigma.\end{aligned} \tag{33}$$

There are two methods to obtain the magnitude of $P_{\alpha\mathrm{APP}}(r,\theta,t)$; either use Eq. (33) directly and obtain an answer involving double integrals or use complex Bessel functions. The direct method involves calculating the magnitude of the product of the beam pattern diffraction terms in Eq. (33),

$$|P_{\alpha\mathrm{APP}}(r,\theta)|=\frac{\rho_0\omega u_0}{2\pi r}e^{-\alpha r}[I_{\alpha C}^2(r,\theta)+I_{\alpha S}^2(r,\theta)]^{1/2}, \tag{34}$$

where

$$\begin{aligned}I_{\alpha C}(r,\theta)&=\int_0^a\int_0^{2\pi}\sigma e^{\alpha[\sigma\sin\theta\cos\varphi-(\sigma^2/2r)]}\\&\quad\times\cos\left[k\left(\sigma\sin\theta\cos\varphi-\frac{\sigma^2}{2r}\right)\right]d\varphi\,d\sigma\end{aligned} \tag{35}$$

and

$$\begin{aligned}I_{\alpha S}(r,\theta)&=\int_0^a\int_0^{2\pi}\sigma e^{\alpha[\sigma\sin\theta\cos\varphi-(\sigma^2/2r)]}\\&\quad\times\sin\left[k\left(\sigma\sin\theta\cos\varphi-\frac{\sigma^2}{2r}\right)\right]d\varphi\,d\sigma.\end{aligned} \tag{36}$$

The complex Bessel function method involves rewriting Eq. (33) in the form

$$\begin{aligned}P_{\alpha\mathrm{APP}}(r,\theta,t)&=i\frac{\rho_0\omega u_0}{2\pi r}e^{i(\omega t-kr)}e^{-\alpha r}\\&\quad\times\int_0^a\sigma e^{-i(k\sigma^2/2r)}e^{-(\alpha\sigma^2/2r)}\\&\quad\times\int_0^{2\pi}e^{i(k-i\alpha)\sigma\sin\theta\cos\varphi}\,d\varphi\,d\sigma\end{aligned} \tag{37}$$

and using Eq. (12) to obtain

$$\begin{aligned}P_{\alpha\mathrm{APP}}(r,\theta,t)&=i\frac{\rho_0\omega u_0}{r}e^{i(\omega t-kr)}e^{-\alpha r}\\&\quad\times\int_0^a\sigma e^{-i(k\sigma^2/2r)}e^{-(\alpha\sigma^2/2r)}\\&\quad\times J_0(\sigma\sin\theta(k-i\alpha))d\sigma.\end{aligned} \tag{38}$$

Formulas for computing Bessel functions with complex arguments are derived in Appendix C where it is shown [Eq. (C4)] that the zeroth order complex Bessel function in Eq. (38) may be represented by a real part, $U_0(\rho,\phi)$, and an imaginary part, $V_0(\rho,\phi)$. Substituting Eq. (C4) into Eq. (38) and calculating the magnitude of the product of the beam pattern diffraction terms

$$|P_{\alpha\mathrm{APP}}(r,\theta)| = \frac{\rho_0 \omega u_0}{r} e^{-\alpha r}[I^2_{\alpha\mathrm{RE}}(r,\theta) + I^2_{\alpha\mathrm{IM}}(r,\theta)]^{1/2}, \tag{39}$$

where

$$I_{\alpha\mathrm{RE}}(r,\theta) = \int_0^a \sigma e^{-(\alpha\sigma^2/2r)} \left[U_0(\rho(\theta),\phi)\sin\left(\frac{k\sigma^2}{2r}\right) - V_0(\rho(\theta),\phi)\cos\left(\frac{k\sigma^2}{2r}\right)\right] d\sigma \tag{40}$$

and

$$I_{\alpha\mathrm{IM}}(r,\theta) = \int_0^a \sigma e^{-(\alpha\sigma^2/2r)} \left[U_0(\rho(\theta),\phi)\cos\left(\frac{k\sigma^2}{2r}\right) + V_0(\rho(\theta),\phi)\sin\left(\frac{k\sigma^2}{2r}\right)\right] d\sigma. \tag{41}$$

Lateral beam profile plots were obtained using both these methods [Eqs. (34) and (39)] and, in general, the results agreed to four significant figures (one part in 1000). Although the complex Bessel function formulas are tedious to program, computations using them executed roughly 15 times faster than the double integral direct method.

To test the accuracy of the attenuated complex Bessel function approximate expression [Eq. (39)] the approximate axial pressure was obtained by setting $r=z$ and $\theta=0$ and integrating by parts to obtain

$$|P_{\alpha\mathrm{APP}}(z,0)| = \frac{\rho_0 \omega u_0}{\sqrt{k^2+\alpha^2}} e^{-\alpha z}\left\{ 1 + e^{(-\alpha a^2/z)} - 2e^{(-\alpha a^2/2z)}\cos\left(\frac{ka^2}{2z}\right)\right\}^{1/2}. \tag{42}$$

This formula also may be obtained from the double integral method. It results, as well, from Eq. (26) after a binomial series expansion with $z>a$. When plots comparing Eq. (42) with the exact axial attenuated pressure [Eq. (26)] were performed, Eq. (42) gave almost identical results to Eq. (26) except at shallow axial distances. In terms of accuracy these plots, not shown here, resembled the real-space axial plots in Fig. 2.

Since increased attenuation shifts the attenuated beam pattern amplitude to lower depths and the accuracy of Eqs. (34) and (39) deteriorates at shallow axial depths, these equations become less useful as the attenuation becomes greater. For more accurate calculations at closer depths the exact attenuated expressions in Appendix D should be used.

At any depth the attenuated unfocused beam pattern will be much lower in magnitude than the unattenuated unfocused beam pattern. But will the shape of the lateral beam profile change with attenuation? To answer this question lateral beam profile calculations were performed for $f=3.5$ MHz, $a=1$ cm, and $c=0.154$ cm/μs at the depth $0.25Y_0=5.692$ cm. The calculations were performed with attenuations η_0 of 0, 0.014, 1.41, and 10 dB/(cm MHz). Figure 8 shows the results of these calculations with all peak pressures normalized to unity.

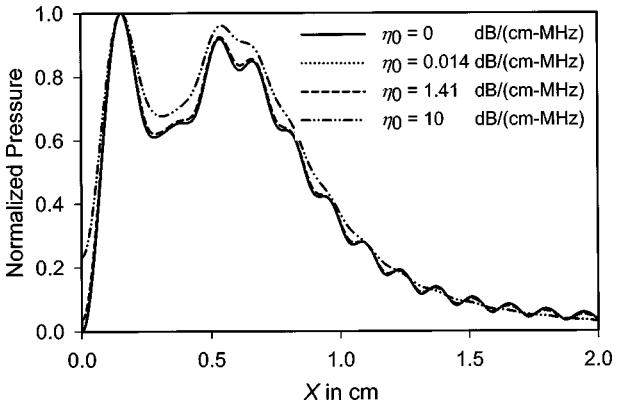


FIG. 8. Change of near zone lateral beam profile with attenuation. The lateral beam profile is plotted at various values of η_0 for $f=3.5$ MHz, $a=1$ cm, and $c=0.154$ cm/μs at a depth of $0.25Y_0=5.682$ cm. The peak pressures are normalized to unity. At low values of attenuation the lateral beam profile is unchanged. At slightly higher values it begins to change with the most notable feature being the increase of the axial minimum from zero. At the highest value of attenuation plotted the lateral beam pattern has smoothed out considerably.

The 0.014 dB/(cm MHz) lateral beam profile is identical with the unattenuated lateral beam profile. The 1.41 dB/(cm MHz) lateral beam profile is slightly different from the unattenuated beam profile. The lateral beam profile oscillations are smoothed out at 10 dB/(cm MHz). At $z=0.25Y_0$ the axial magnitude of the lateral beam profile should be close to zero because an axial minimum occurs here. But in Fig. 8 it is seen that the axial magnitude of the lateral beam profile increases from zero with increasing attenuation. From Figs. 6(b) and 7 these nonzero minima are seen to be due only to changes in the lateral beam profile with attenuation and not axial shifts in these minima.

This was further verified by computations of the attenuated exact axial pressure [Eq. (26)] and the attenuated complex Bessel approximate axial pressure [Eq. (42)] at this depth. The results of these computations showed that the axial minimum had a negligible shift and was greater than zero when $\eta_0=1.41$ dB/(cm MHz). At $\eta_0=10$ dB/(cm MHz) the minimum was gone and the axial pressure had a decaying exponential like appearance.

Beam pattern shifting toward the source due to attenuation is greatest at the broad axial maximum at Y_0 (cf. Fig. 7). Figure 9 shows the change in the lateral beam profile shape an experimenter would measure if he or she first located the shifted Y_0 at each fluid attenuation value with an axial measurement and then measured the lateral beam profile at this depth. Even for the moderate attenuation values used in Fig. 9, the ''tracked'' lateral beam profile rapidly develops sidelobes larger in magnitude than the axial peak.

Lateral profile plots at many depths with various values of the source parameters and attenuation were calculated for the approximate beam pattern pressure equations derived

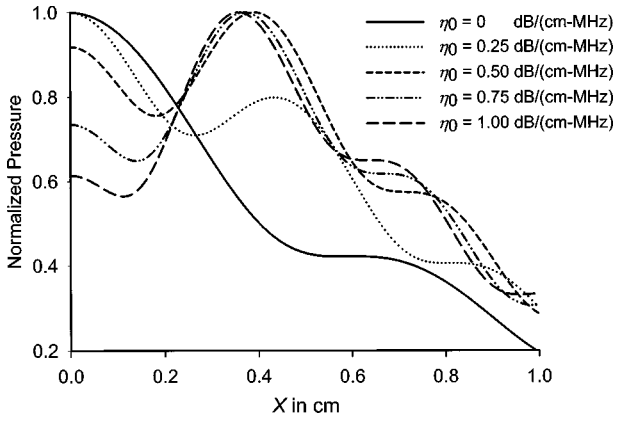


FIG. 9. Last axial maximum lateral beam profile at various attenuations. The lateral profiles at Y_0, located experimentally at each attenuation value by an axial measurement, are normalized to unity at their peaks. For $f=3.5$ MHz, $a=1$ cm, and $c=0.154$ cm/μs the last axial maximum is at 22.727 cm for $\eta_0=0$. For $\eta_0=0.25$ dB/(cm MHz) it is at 16.161 cm. For $\eta_0=0.50$ dB/(cm MHz) it is at 14.571 cm. For $\eta_0=0.75$ dB/(cm MHz) it is at 13.789 cm. And when $\eta_0=1.00$ dB/(cm MHz) it is at 13.310 cm. As the attenuation increases side lobes appear which are larger in magnitude than the axial peak.

here and the exact equations in Appendix D. At deep axial depths there was excellent agreement. At shallow axial depths for moderate values of attenuation the level of agreement was the same as between the unattenuated approximate and exact equations studied above. At shallow depths and high values of attenuation the agreement between the approximate and exact equations improved.

IV. SUMMARY

The single integral approximate formulas for the beam pattern of a nonattenuated circular unfocused flat piston source in fluids derived here have been found to be accurate at shallow beam depths. Comparison of axial approximate and exact pressure equations indicate the depth vs accuracy relation. High accuracy at short depths may be attained by using the normalized form of the nonattenuated beam pattern.

Since fluid attenuation measurements routinely are performed along the beam axis and a plane wave attenuation factor is assumed for data reduction, the results of this analysis should aid experimenters in avoiding inaccurate attenuation measurement results when using unfocused plane piston sources. The single integral equations derived here permit rapid, accurate computation of the attenuated circular unfocused flat piston beam pattern.

APPENDIX A: JINC FUNCTION APPROXIMATION

A Jinc function directivity term will approximate the lateral beam profile when the $I_S(r,\theta)$ integral is negligible compared to the $I_C(r,\theta)$ integral [Eqs. (15) and (16)]. For an order of magnitude computation assume that the ratio of these integrals is equal to the ratio of the average values of the sine and cosine terms in their kernels. The integration variable σ has an average value

$$\frac{1}{a}\int_0^a \sigma\, d\sigma = \frac{a}{2} \tag{A1}$$

neglecting the other terms in the kernel. Using this average value of σ along with relation $r=nY_0$, the arguments of the sine and cosine terms become

$$\frac{k\sigma^2}{2r} \approx \frac{\pi}{4n}. \tag{A2}$$

The larger n (and r) the smaller in magnitude the argument of the average values of the sine and cosine terms and the more negligible the $I_s(r,\theta)$ integral compared to the $I_c(r,\theta)$ integral. At what value of n can the $I_s(r,\theta)$ integral be ignored? Take the onset of the far field when $n=6.41$, the ratio of the sine and cosine is

$$\frac{\sin 7^\circ}{\cos 7^\circ} = \frac{0.1222}{0.9925} = 0.123. \tag{A3}$$

However these two integrals combine as the square root of the sum of their squares in Eq. (14), so

$$\begin{aligned}\sqrt{\cos^2 7^\circ + \sin^2 7^\circ} &= \cos 7^\circ \sqrt{1+\frac{\sin^2 7^\circ}{\cos^2 7^\circ}} \\ &= 0.9925\sqrt{1+.015} \\ &\cong 0.9925(1+0.0075).\end{aligned} \tag{A4}$$

Thus, the sine integral will have less than a 1% contribution in Eq. (14) and can be neglected when $n>6.41$.

APPENDIX B: ATTENUATION COEFFICIENTS

For both plane and spherical ultrasound waves the attenuation term in Eq. (23) produces a multiplicative attenuation factor of the form

$$e^{-\alpha\cdot \text{distance}}. \tag{B1}$$

The attenuation coefficient α has the units cm^{-1} and the general frequency dependence

$$\alpha = f^{\text{m}}\alpha_0, \tag{B2}$$

where f is the frequency in MHz and α_0 has the units $\text{Np/(cm MHz}^m)$.

The attenuation coefficient also can be expressed in dB/cm

$$\eta = f^m \eta_0, \tag{B3}$$

where η_0 has the units $\text{dB/(cm MHz}^m)$. This form is more useful when considering propagation in human tissue; which, like a fluid, only supports longitudinal waves. Using the relation

$$\eta = 8.686\alpha, \tag{B4}$$

for propagation in tissue

$$\alpha = \frac{f^m}{8.686}\eta_0. \tag{B5}$$

For propagation in tissue m is a little greater than unity and for propagation in most liquids $m=2$. For convenience

in calculations, when considering tissue propagation Eq. (B5) should be used for α with $m=1$ and when considering propagation in liquids Eq. (B2) should be used for α with $m=2$. In many tables of liquid acoustic properties the attenuation, α/f^2, is presented in the units $\alpha/f^2\times10^{-17}\,\mathrm{s}^2/\mathrm{cm}$. Multiplying by 10^{12} will change it to the more convenient units of $\mathrm{Np}/(\mathrm{cm\,MHz}^2)$.

APPENDIX C: BESSEL FUNCTIONS WITH COMPLEX ARGUMENTS

The computation begins with the Bessel function series expansion[10]

$$J_n(s)=\sum_{m=0}^{\infty}\frac{(-1)^m}{m!(n+m)!}\left(\frac{1}{2}s\right)^{n+2m}. \tag{C1}$$

s is a complex number that will be expressed in the general complex polar form

$$s=\rho e^{i\phi}, \tag{C2}$$

where ρ and ϕ are real numbers. Only the J_0 complex Bessel function will be considered here. From Eqs. (C1) and (C2)

$$J_0(\rho e^{i\phi})=\sum_{m=0}^{\infty}\frac{(-1)^m}{(m!)^2}\left(\frac{\rho}{2}\right)^{2m}e^{i2m\phi}. \tag{C3}$$

Using Euler's formula

$$e^{ix}=\cos x+i\sin x$$

this complex Bessel function expansion can be grouped into real (U) and imaginary (V) parts[11]

$$J_0(\rho e^{i\phi})=U_0(\rho e^{i\phi})+iV_0(\rho e^{i\phi}), \tag{C4}$$

where

$$U_0(\rho e^{i\phi})=\sum_{m=0}^{\infty}\frac{(-1)^m}{(m!)^2}\left(\frac{\rho}{2}\right)^{2m}\cos 2m\phi \tag{C5}$$

and

$$V_0(\rho e^{i\phi})=\sum_{m=0}^{\infty}\frac{(-1)^m}{(m!)^2}\left(\frac{\rho}{2}\right)^{2m}\sin 2m\phi. \tag{C6}$$

The Bessel function complex argument in attenuated beam pattern calculations for a circular aperture unfocused piston source is [Eq. (38)]

$$\rho e^{i\phi}=k_c\sigma\sin\theta=\sigma\sin\theta\left(\frac{2\pi f}{c}-i\alpha\right). \tag{C7}$$

The complex plane polar angle ϕ is

$$\phi=\tan^{-1}\left(-\frac{c}{2\pi f}\alpha\right). \tag{C8}$$

For propagation in tissue the appropriate relation is [Eq. (B5)]

$$\phi=\tan^{-1}\left(-\frac{c}{54.58}\eta_0\right). \tag{C9}$$

For propagation in most liquids the appropriate relation is [Eq. (B2)]

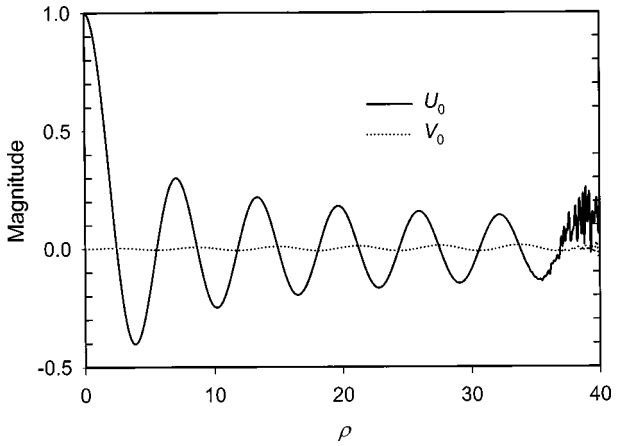


FIG. 10. The real and imaginary parts of the complex J_0 for average human tissue where $\phi=-0.162°$.

$$\phi=\tan^{-1}\left(-\frac{cf}{2\pi}\alpha_0\right), \tag{C10}$$

where c is in units of cm/μs and f is in units of MHz.

Once ϕ is known the modulus ρ is given by

$$\rho(\theta)=\frac{2\pi f\sigma\sin\theta}{c\cos\phi} \tag{C11}$$

or

$$\rho(x)=\frac{2\pi f\sigma x}{c\cos\phi\sqrt{z^2+x^2}}. \tag{C12}$$

For average human tissue $\eta_0=1$ dB/(cm MHz) and Eq. (C9) yields $\phi=-0.162°$. Figure 10 demonstrates $U_0(\rho,\phi)$ and $V_0(\rho,\phi)$ in this case as ρ varies from 0 to 40. At larger values of ρ the series expansions become unstable because the series terms are oscillating between ever greater positive and negative numbers. It is clearly seen that the imaginary part of the complex Bessel function cannot be ignored.

Computing $U_0(\rho,\phi)$ and $V_0(\rho,\phi)$ for large arguments requires the asymptotic series expansion[12]

$$J_0(s)=\left(\frac{2}{\pi s}\right)^{1/2}\left[P_0(s)\cos\left(s-\frac{\pi}{4}\right)-Q_0(s)\sin\left(s-\frac{\pi}{4}\right)\right], \tag{C13}$$

where

$$P_0(s)\approx 1-\frac{1^2 3^2}{2!(8s)^2}+\frac{1^2 3^2 5^2 7^2}{4!(8s)^4}-\frac{1^2 3^2 5^2 7^2 9^2 11^2}{6!(8s)^6}+\cdots \tag{C14}$$

and

$$Q_0(s)\approx-\frac{1^2}{1!8s}+\frac{1^2 3^2 5^2}{3!(8s)^3}-\frac{1^2 3^2 5^2 7^2 9^2}{5!(8s)^5}+\cdots. \tag{C15}$$

Substituting Eq. (C2) into the asymptotic series expansion and simplifying

$$U_0(\rho,\phi)\cong A(\rho,\phi)E(\rho,\phi)C(\rho,\phi)-A(\rho,\phi)F(\rho,\phi)D(\rho,\phi)-A(\rho,\phi)G(\rho,\phi)I(\rho,\phi)+A(\rho,\phi)H(\rho,\phi)K(\rho,\phi)$$
$$+B(\rho,\phi)C(\rho,\phi)F(\rho,\phi)+B(\rho,\phi)E(\rho,\phi)D(\rho,\phi)-B(\rho,\phi)H(\rho,\phi)I(\rho,\phi)-B(\rho,\phi)G(\rho,\phi)K(\rho,\phi), \tag{C16}$$

$$V_0(\rho,\phi)\cong A(\rho,\phi)F(\rho,\phi)C(\rho,\phi)+A(\rho,\phi)E(\rho,\phi)D(\rho,\phi)-A(\rho,\phi)H(\rho,\phi)I(\rho,\phi)-A(\rho,\phi)G(\rho,\phi)K(\rho,\phi)$$
$$-B(\rho,\phi)E(\rho,\phi)C(\rho,\phi)+B(\rho,\phi)F(\rho,\phi)D(\rho,\phi)+B(\rho,\phi)G(\rho,\phi)I(\rho,\phi)-B(\rho,\phi)H(\rho,\phi)K(\rho,\phi), \tag{C17}$$

where

$$A(\rho,\phi)=\left(\frac{2}{\pi\rho}\right)^{1/2}\cos\left(\frac{\phi}{2}\right),$$
$$B(\rho,\phi)=\left(\frac{2}{\pi\rho}\right)^{1/2}\sin\left(\frac{\phi}{2}\right), \tag{C18}$$

$$C(\rho,\phi)=\frac{1}{\sqrt{2}}\cosh(\rho\sin\phi)[\cos(\rho\cos\phi)+\sin(\rho\cos\phi)], \tag{C19}$$

$$D(\rho,\phi)=\frac{1}{\sqrt{2}}\sinh(\rho\sin\phi)[\cos(\rho\cos\phi)-\sin(\rho\cos\phi)], \tag{C20}$$

$$E(\rho,\phi)=1-\frac{0.0703}{\rho^2}\cos 2\phi+\frac{0.1122}{\rho^4}\cos 4\phi-\frac{0.5725}{\rho^6}\cos 6\phi, \tag{C21}$$

$$F(\rho,\phi)=\frac{0.0703}{\rho^2}\sin 2\phi-\frac{0.1122}{\rho^4}\sin 4\phi+\frac{0.5725}{\rho^6}\sin 6\phi, \tag{C22}$$

$$G(\rho,\phi)=-\frac{0.125}{\rho}\cos\phi+\frac{0.0732}{\rho^3}\cos 3\phi-\frac{0.2271}{\rho^5}\cos 5\phi, \tag{C23}$$

$$H(\rho,\phi)=\frac{0.125}{\rho}\sin\phi-\frac{0.0732}{\rho^3}\sin 3\phi+\frac{0.2271}{\rho^5}\sin 5\phi, \tag{C24}$$

$$I(\rho,\phi)=\frac{1}{\sqrt{2}}\cosh(\rho\sin\phi)[\sin(\rho\cos\phi)-\cos(\rho\cos\phi)], \tag{C25}$$

and

$$K(\rho,\phi)=\frac{1}{\sqrt{2}}\sinh(\rho\sin\phi)[\cos(\rho\cos\phi)+\sin(\rho\cos\phi)]. \tag{C26}$$

The predictions of Eqs. (C5) and (C16) for $U_0(\rho,\phi)$ and (C6) and (C17) for $V_0(\rho,\phi)$ were compared and excellent agreement was found. For $\rho>2$ [and up to the highest ρ values possible with Eqs. (C5) and (C6)] the agreement was to four significant figures except for the very lowest magnitudes of $U_0(\rho,\phi)$ and $V_0(\rho,\phi)$ where the agreement was a one to three unit variation in the fourth significant figure (about 3 parts per 1000). This was probably caused by the choice of using only four digit numbers in Eqs. (C21)–(C24). So the approximate signs in Eqs. (C16) and (C17) effectively may be replaced by equals signs.

For numerical computations Eqs. (C5) or (C6) can be combined with Eqs. (C16) or (C17) to represent $U_0(\rho,\phi)$ or $V_0(\rho,\phi)$ for all values of ρ. A conservative combination was utilized here using Eqs. (C5) and (C6) up to $\rho=5$, summing over 15 terms, and Eqs. (C16) and (C17) for $\rho>5$.

APPENDIX D: EXACT PRESSURE EQUATIONS

An expression for the exact attenuated pressure is obtained by substituting Eq. (23) into Eq. (1) and using Eq. (9),

$$P_{\alpha\mathrm{ex}}(r,\theta,t)=i\frac{\rho_0\omega u_0}{2\pi}e^{i\omega t}\int_0^a\int_0^{2\pi}e^{-\alpha\sqrt{r^2+\sigma^2-2r\sigma\sin\theta\cos\varphi}}\times\frac{\sigma e^{-ik(\sqrt{r^2+\sigma^2-2r\sigma\sin\theta\cos\varphi})}}{\sqrt{r^2+\sigma^2-2r\sigma\sin\theta\cos\varphi}}d\varphi\,d\sigma. \tag{D1}$$

This may be rewritten as

$$P_{\alpha\mathrm{ex}}(r,\theta,t)=i\frac{\rho_0\omega u_0}{2\pi}e^{i(\omega t-kr)}\int_0^a\int_0^{2\pi}e^{-\alpha\sqrt{r^2+\sigma^2-2r\sigma\sin\theta\cos\varphi}}\times\frac{\sigma e^{-ik(\sqrt{r^2+\sigma^2-2r\sigma\sin\theta\cos\varphi}-r)}}{\sqrt{r^2+\sigma^2-2r\sigma\sin\theta\cos\varphi}}d\varphi\,d\sigma. \tag{D2}$$

The magnitude of $P_{\alpha\mathrm{ex}}$ is computed, at any time t, as the magnitude of the product of the beam pattern diffraction terms of Eq. (D2)

$$|P_{\alpha\mathrm{ex}}(r,\theta)|=\frac{\rho_0\omega u_0}{2\pi}\sqrt{I^2_{\alpha\mathrm{RE}}(r,\theta+I^2_{\alpha\mathrm{IM}}(r,\theta)}, \tag{D3}$$

where

$$I_{\alpha \mathrm{RE}}(r,\theta)=\int_0^a\int_0^{2\pi} e^{-\alpha\sqrt{r^2+\sigma^2-2r\sigma\sin\theta\cos\varphi}} \times\frac{\sigma\sin(k(\sqrt{r^2+\sigma^2-2r\sigma\sin\theta\cos\varphi}-r))}{\sqrt{r^2+\sigma^2-2r\sigma\sin\theta\cos\varphi}} \times d\varphi\, d\sigma \tag{D4}$$

and

$$I_{\alpha \mathrm{IM}}(r,\theta)=\int_0^a\int_0^{2\pi} e^{-\alpha\sqrt{r^2+\sigma^2-2r\sigma\sin\theta\cos\varphi}} \times\frac{\sigma\cos(k(\sqrt{r^2+\sigma^2-2r\sigma\sin\theta\cos\varphi}-r))}{\sqrt{r^2+\sigma^2-2r\sigma\sin\theta\cos\varphi}} \times d\varphi\, d\sigma. \tag{D5}$$

[1] J. Zemanek, "Beam behavior within the nearfield of a vibrating piston," J. Acoust. Soc. Am. **49**, 181–191 (1971).

[2] J. C. Lockwood and J. G. Willette, "High-speed method for computing the exact solution for the pressure variations in the nearfield of a baffled piston," J. Acoust. Soc. Am. **53**, 735–741 (1973).

[3] H. Seki, A. Granato, and R. Truell, "Diffraction effects in the ultrasonic field of a piston source and their importance in the accurate measurement of attenuation," J. Acoust. Soc. Am. **28**, 230–238 (1956).

[4] K. Beissner, "Exact integral expression for the diffraction loss of a circular piston source," Acustica **49**, 212–217 (1981).

[5] A. Goldstein and R. L. Powis, "Medical ultrasonic diagnostics," in *Ultrasonic Instruments and Devices I: Reference for Modern Instrumentation, Techniques and Technology*, edited by E. P. Papadakis, Volume 23 in *Physical Acoustics Series*, edited by R. N. Thurston and A. D. Pierce (Academic, New York, 1998), p. 49.

[6] A. Goldstein, D. R. Gandhi, and W. D. O'Brien, Jr., "Diffraction effects in hydrophone measurements," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 972–979 (1998).

[7] D. Cathignol, O. A. Sapozhnikov, and J. Zhang, "Lamb waves in piezoelectric focused radiator as a reason for discrepancy between O'Neil's formula and experiment," J. Acoust. Soc. Am. **101**, 1286–1297 (1997).

[8] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982), pp. 176–182.

[9] L. E. Kinsler and A. R. Frey, *Fundamentals of Acoustics*, 2nd ed. (Wiley, New York, 1962), pp. 166–169.

[10] G. N. Watson, *Theory of Bessel Functions*, 2nd ed. (Cambridge University Press, London, 1944), p. 15.

[11] *Table of the Bessel Functions* $J_0(z)$ *and* $J_1(z)$ *for Complex Arguments*, 2nd ed. prepared by the Mathematical Tables Project National Bureau of Standards (Columbia University Press, New York, 1947).

[12] H. B. Dwight, *Tables of Integrals and Other Mathematical Data*, 4th ed. (MacMillan, New York, 1961), p. 192.

Spontaneous pattern formation in an acoustical resonator

V. J. Sánchez-Morcillo[a)]
Departamento de Física Aplicada, Universidad Politécnica de València, Carretera Nazaret-Oliva S/N, 46730 Grao de Gandia, Spain

(Received 22 May 2003; revised 29 October 2003; accepted 31 October 2003)

A dynamical system of equations describing parametric sound generation (PSG) in a dispersive large aspect ratio resonator is derived. The model generalizes previously proposed descriptions of PSG by including diffraction effects and is analogous to the model used in theoretical studies of optical parametric oscillation. A linear stability analysis of the solution below the threshold of subharmonic generation reveals the existence of a pattern forming instability, which is confirmed by numerical integration. The conditions of emergence of periodic patterns in transverse space are discussed in the acoustical context. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1635416]

PACS numbers: 43.25.Gf, 43.25.Ts [MFH] Pages: 111–119

I. INTRODUCTION

In recent years, pattern formation in systems driven far from equilibrium has became an active field of research in many areas of nonlinear science. Apart from peculiarities of particular systems, an outstanding property of pattern formation is its universality, evidenced when the dynamical models describing the different phenomena (either hydrodynamical, chemical, optical, or others) can be reduced, under several approximations, to the same order parameter equation. These equations are few and well known, such as Ginzburg–Landau or Swift Hohenberg.[1]

A key concept in pattern forming systems is the aspect ratio. When the evolution of the variables is restricted to a bounded region of space, or cavity, the aspect ratio is defined as the ratio of transverse to longitudinal sizes of the cavity. In hydrodynamical Rayleigh–Bènard convection, for example, the aspect ratio is determined by the ratio between the height of the fluid and the area of the container. In problems of nonlinear wave interaction, this parameter is related to the Fresnel number, usually defined as

$$F=\frac{a^2}{\lambda L}, \tag{1}$$

where a is the characteristic transverse size of the cavity (for example, the area of a plane radiator), λ is the wavelength, and L is the length of the cavity in the direction of propagation, considered the longitudinal axis of the cavity.

Spontaneous pattern formation is observed in large aspect ratio nonlinear systems driven by an external input, where the possibility of excitation of many transverse modes (a continuum for an infinite transverse dimension) is considered. When the amplitude of the external input reaches a critical threshold value, large enough to overcome the losses produced by dissipative processes in the system, a symmetry breaking transition occurs, carrying the system from an initially homogeneous to a inhomogeneous state, usually with spatial periodicity. These solutions have been often called dissipative structures.[2] A paradigmatic example in pattern formation studies has been the Rayleigh–Bènard convection in a fluid layer heated from below, where roll or hexagonal patterns are excited above a given temperature threshold.[1]

This scenario differs with that observed in small aspect ratio systems such as, for example, a waveguide resonator of finite cross section. In this case, boundary-induced spatial patterns are selected not by the nonlinear properties of the medium inside the cavity, but by the transverse boundary conditions, and correspond to the excitation of one or few transverse modes of the cavity.[3] In this sense, boundary-induced patterns are of linear nature.

Guided by the analogies with other physical systems, we can expect that an ideal system for such effects to be observed in acoustics consists in a resonator of plane walls (acoustical interferometer) with infinite transverse size. In practice, the large aspect ratio condition could be fulfilled if the pumped area is finite, but large in comparison with the spatial scale imposed by the cavity length and the field wavelength, as follows from (1).

Parametrically driven systems offer many examples of spontaneous pattern formation. For example, parametric excitation of surface waves by a vertical excitation (Faraday instability),[4] vibrated granular layers,[5] spin waves in ferrites and ferromagnets and Langmuir waves in plasmas parametrically driven by a microwave field,[6] or the optical parametric oscillator[7,8] have been studied.

The behavior of nonlinear waves in large aspect ratio resonators has been extensively studied in nonlinear optics (for a recent review, see Ref. 9), where a rich variety of patterns has been observed. On the other side, it is well known that optical and acoustical waves share many common phenomena, under restricted validity conditions.[10]

In nonlinear acoustics, a phenomenon which belongs to the class of the previous examples is the parametric sound amplification. It consists in the resonant interaction of a triad of waves with frequencies ω_0, ω_1, and ω_2, for which the following energy and momentum conservation conditions are fulfilled:

a)Electronic mail: victorsm@fis.upv.es

$$\omega_0 = \omega_1 + \omega_2,$$
$$\vec{k}_0 = \vec{k}_1 + \vec{k}_2 + \Delta\vec{k}, \tag{2}$$

where $\Delta\vec{k}$ is a small phase mismatch. The process is initiated by an input pumping wave of frequency ω_0 which, due to the propagation in the nonlinear medium, generates a pair of waves with frequencies ω_1 and ω_2. When the wave interaction occurs in a resonator, a threshold value for the input amplitude is required, and the process is called parametric sound generation. In acoustics, this process has been described before by several authors under different conditions, either theoretical and experimental. In Refs. 11–13, the one dimensional case (colinearly propagating waves) is considered. In Ref. 3, the problem of interaction between concrete resonator modes, with a given transverse structure, is studied. In both cases, small aspect ratio resonators containing liquid and gas respectively are considered. More recently, parametric interaction in a large aspect ratio resonator filled with superfluid He^4 has been investigated.[14,15]

The phenomenon of parametric sound generation is analogous to the optical parametric oscillation in nonlinear optics. However, an important difference between acoustics and optics is the absence of dispersion in the former. Dispersion, which makes the phase velocity of the waves to be dependent on its frequency, allows that only few waves, those satisfying given synchronism conditions, participate in the process.

In a nondispersive medium, all the harmonics of each initial monochromatic wave propagate synchronously. As a consequence, the spectrum broadens during propagation and the energy is continuously pumped into the higher harmonics, which eventually leads to shock formation.

Optical media are in general dispersive, but acoustical media are not. In finite geometries, such as waveguides[16] or resonators,[17] the dispersion is introduced by the boundaries. Different cavity modes propagate at different angles, and then with different ''effective'' phase velocities. However, in unbounded systems boundary-induced dispersion is not present.

Other dispersion mechanisms have been proposed in nonlinear acoustics, such as bubbly media or layered (periodic) media.[18] In all this systems, dispersion appears due to the introduction of additional spatial or temporal scales in the system, which makes sound velocity propagation to be wavelength dependent. Other proposed mechanisms are, for example, media with selective absorption, in which selected spectral components experience strong losses and may removed from the wave field.[19]

Pattern formation in acoustics has been reported previously in the context of acoustic cavitation,[20,21] where the coupling between the sound field amplitude and the bubble distribution is considered. We do not consider here the dynamics of the medium, which is assumed to be at rest, but the coupling between different frequency components.

The aim of the paper is twofold. On one side, a rigurous derivation of the dynamical model describing the parametric interaction of acoustic waves in a large aspect ratio cavity is presented. The derived model is analogous to the system of equations describing parametric oscillation in an optical resonator, and consequently their solutions are known. On the other side, among other peculiarities, the model predicts a pattern forming instability, which is confirmed by a numerical integration. We review these properties in the acoustical context, giving some estimations of the acoustical parameters which can make the model closer to real operating conditions.

II. MODEL EQUATIONS

The starting point of the theoretical analysis is the nonlinear wave equation, which written in terms of the pressure p reads[18]

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} + \frac{\delta}{c^4}\frac{\partial^3 p}{\partial t^3} = -\frac{\beta}{\rho_0 c^4}\frac{\partial^2 p^2}{\partial t^2} - \left(\nabla^2 + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\mathcal{L}, \tag{3}$$

where $\mathcal{L} = \frac{1}{2}\rho_0 v^2 - p^2/2\rho_0 c^2$ is the Lagrangian density, with ρ_0 the ambient density and v the particle velocity, δ is the diffusivity of sound, defined as[18]

$$\delta = \frac{1}{\rho_0}\left(\frac{4}{3}\mu + \mu_B + \kappa\left(\frac{1}{c_v} - \frac{1}{c_p}\right)\right), \tag{4}$$

and $\beta = 1 + B/2A$ is the nonlinearity coefficient.

In the case of colinearly propagating plane waves, the Lagrangian density term vanishes, since the linear impedance relation $p = \rho_0 c v$ holds. However, even in the case of slightly diverging waves (plane waves propagating at a small angle), the last term in Eq. (3) is much smaller in magnitude than the other terms in the equation, and in this approximation its effects can be ignored.[16]

Furthermore, in this case the nonlinearity parameter can be considered to be independent of the interaction angle. It has been shown[16] that, for the process described by Eqs. (2), when the waves with frequencies ω_1 and ω_2 propagate at an angle θ, the nonlinearity parameter for the ω_0 wave is expressed as

$$\beta_0(\theta) = \cos\theta + 4\frac{\omega_1\omega_2}{\omega_0^2}\sin^4\left(\frac{\theta}{2}\right) + \frac{B}{2A}, \tag{5}$$

which reduces to $\beta = 1 + B/2A$ when θ is small.

Under this assumption, the field distribution can be properly described by the wave equation

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} + \frac{\delta}{c^4}\frac{\partial^3 p}{\partial t^3} = -\frac{\beta}{\rho_0 c^4}\frac{\partial^2 p^2}{\partial t^2}, \tag{6}$$

which is the well known Westervelt equation.[22]

Equation (6) describes the propagation of waves in a nondispersive homogeneous medium, but also in a medium that possess some dispersion mechanism, such as a bubbly liquid when the field frequencies are much lower than the resonance frequency of the bubbles.[23,24] It must be noted that, in a dispersive medium, the nonlinearity parameter β depends on the dispersion mechanism.

We consider in the following that the wave interaction is only effective among three resonant frequencies, for which

the relations (2) hold. Taking into account that, due to reflections in the walls, there exist waves propagating simultaneously in opposite directions, the field inside the resonator can then be expanded as

$$p(\mathbf{r},t)=\sum_{j=1}^{3} P_j(\mathbf{r},t), \tag{7}$$

where $P_j(\mathbf{r},t)$ are the wave components related with the frequency ω_j, given by

$$P_j(\mathbf{r},t)=p_j(x,y,t)\cos(k_j^c z)e^{-i\omega_j t}+\text{c.c.}, \tag{8}$$

where $k_j^c=m\pi/L$ is a cavity eigenmode.

A more general solution can be proposed, consisting in a superposition of quasi-planar waves, the field being decomposed in forward (F) and backward (B) waves, respectively. The quasi-planar assumption implies that the amplitudes may depend on the longitudinal coordinate z, representing a slow evolution compared with the scale given by k_j. The field at frequency ω_j is then expressed as

$$P_j(\mathbf{r},t)=[F_j(\mathbf{r},t)e^{ik_j z}+B_j(\mathbf{r},t)e^{-ik_j z}]e^{-i\omega_j t}+\text{c.c.}, \tag{9}$$

where $\mathbf{r}=(x,y,z)$.

The case $\omega_0=2\omega_1$ corresponds to degenerate interaction ($\omega_1=\omega_2=\omega_0/2$), and describes the process of second harmonic generation or subharmonic parametric generation, depending on whether the pumping wave oscillates at ω_0 or ω_1. In the following, the degenerate interaction case is considered, where

$$p(\mathbf{r},t)=P_0(\mathbf{r},t)+P_1(\mathbf{r},t) \tag{10}$$

with the amplitudes given in Eq. (9).

Substituting Eq. (10) in Eq. (6), and projecting the resulting equation on each of the mode frequencies, two coupled wave equations are found,

$$\nabla^2 P_j-\frac{1}{c_j^2}\frac{\partial^2 P_j}{\partial t^2}+\frac{\delta_j}{c_j^4}\frac{\partial^3 P_j}{\partial t^3}=-\frac{\beta}{\rho_0 c_j^4}\frac{\partial^2 (p^2)_j}{\partial t^2}, \tag{11}$$

where $j=0$, 1 and $c_j=\omega_j/k_j=c(\omega_j)$ is the phase velocity of the waves.

The evolution equations for the amplitudes p_j at the different frequencies can derived be from Eq. (11) under several approximations. The complete derivation is given in the Appendix. If we assume that (i) the amplitudes are slowly varying in z and t, (ii) the reflectivity $\mathcal{R}_j$ at the boundaries is high, (iii) the field frequencies ω_j are close to one resonator frequency, ω_j^c, and that (iv) only one longitudinal mode is excited by each frequency component, the evolution is ruled by the following dynamical equations:

$$\begin{aligned}\frac{\partial p_0}{\partial t}&=E-\gamma_0(1+i\Delta_0)p_0+ia_0\nabla_\perp^2 p_0-i\frac{b_0}{4}p_1^2,\\ \frac{\partial p_1}{\partial t}&=-\gamma_1(1+i\Delta_1)p_1+ia_1\nabla_\perp^2 p_1-i\frac{b_1}{2}p_1^* p_0,\end{aligned} \tag{12}$$

where $p_j(x,y,t)$ are the amplitudes corresponding to Eq. (8), the nonlinearity parameter is defined as $b_j=\beta k_j/2\rho_0 c_j$, and the asterisk denoted complex conjugation. The other parameters in Eqs. (12) are the pump E, detuning Δ_j, losses γ_j, and diffraction a_j and are defined as

$$E=\frac{c_0}{2L}\frac{\sqrt{\mathcal{T}_0}}{\sqrt{\mathcal{R}_0}}Ye^{i\theta_0/2}, \tag{13}$$

$$\Delta_j=\frac{\theta_j}{|\ln\mathcal{R}_j|+(2L/c_j)\kappa_j}, \tag{14}$$

$$\gamma_j=\frac{|\ln\mathcal{R}_j|c_j}{2L}+\kappa_j, \tag{15}$$

$$a_j=\frac{c_j}{2k_j}, \tag{16}$$

where Y is the amplitude of the incident wave at frequency ω_0, $\mathcal{R}_j$ are the reflectivities at the boundaries, $\theta_j=2L(\omega_j^c-\omega_j)/c_j$ is proportional to the frequency mismatch, and $\kappa_j=\delta k_j^2/2$ is a loss factor related with the diffusivity of sound. Note that the total loss parameter γ_j takes into account both loss mechanisms, due to absorption and reflectivity at the boundaries (energy leakage of the resonator). Also, from Eqs. (14) and (15) the detuning can be written as $\Delta_j=(\omega_j^c-\omega_j)/\gamma_j$.

Finally, the field amplitudes are normalized to leave the model in its final form,

$$\begin{aligned}\frac{1}{\gamma_0}\frac{\partial A_0}{\partial t}&=\mathcal{E}-(1+i\Delta_0)A_0+ia_0\nabla^2 A_0-A_1^2,\\ \frac{1}{\gamma_1}\frac{\partial A_1}{\partial t}&=-(1+i\Delta_1)A_1+ia_1\nabla^2 A_1+A_0A_1^*,\end{aligned} \tag{17}$$

where $A_0=i(b_1/2\gamma_1)p_0$, $A_1=i\sqrt{b_0 b_1/8\gamma_0\gamma_1}p_1$, and $\mathcal{E}=i(b_1/2\gamma_0\gamma_1)E$.

The system of equations (17), together with their complex conjugates, consists in the model for parametric interaction of acoustic waves in large aspect ratio resonators. Equations (17) are suitable for the description of the spatiotemporal evolution of the pressure envelope waves oscillating at the fundamental (ω_0) and subharmonic (ω_1) frequencies.

These equations have been studied in the context of nonlinear optics, as a model for optical parametric oscillation.[7] In the following we review some of the basic properties of the solutions of Eqs. (17) and their application to the case of the acoustic resonator.

III. HOMOGENEOUS SOLUTIONS

The stationary and spatially homogeneous solutions of Eqs. (17) are obtained when the temporal derivatives and the transverse diffraction term vanish. In this case, the model reduces to the one derived for parametric sound generation in a resonator with rectangular cross section $a\times b$, described in Ref. 3, where its homogeneous solutions were obtained. We next review these solutions in the present notation.

The simplest homogeneous solution corresponds to the trivial solution,

$$\overline{A_0}=\frac{\mathcal{E}}{1+i\Delta_0}, \quad \overline{A_1}=0, \tag{18}$$

characterized by a null value of the subharmonic field inside the resonator. This solution exists for low values of the pump amplitude (below the instability threshold to be discussed in the next section).

For larger pump values, also the subharmonic field has a nonzero amplitude, given by

$$|\overline{A_1}|^2 = -1 + \Delta_0\Delta_1 \pm \sqrt{\mathcal{E}^2 - (\Delta_1 + \Delta_1)^2}, \tag{19}$$

while the stationary fundamental intensity $|A_0|^2$ takes the value

$$|\overline{A_0}|^2 = 1 + \Delta_1^2, \tag{20}$$

which is independent of the value of the injected pump.

The emergence of this finite amplitude solution corresponds to the process of parametric generation. In the frame of the plane wave model it has been shown theoretically, and confirmed experimentally,[3] that the trivial solution (18) bifurcates in the nontrivial one, and the subharmonic field emerges with an amplitude given by Eq. (19). This bifurcation is supercritical when $\Delta_0\Delta_1 < 1$, and subcritical when $\Delta_0\Delta_1 > 1$. In the latter case, both homogeneous solutions can coexist for given sets of the parameters.

In the next section this scenario is generalized by including diffraction effects in the model.

IV. LINEAR STABILITY ANALYSIS

In order to study the stability of the trivial solution (18) against space-dependent perturbations, consider a deviation of this state, given by

$$A_j(x,y,t) = \overline{A_j} + \delta A_j(x,y,t). \tag{21}$$

Assuming that the deviations are small with respect to the stationary values, after substitution of Eq. (21) in (17) the resulting system can be linearized in the perturbations δA_j. This leads to the linear system of equations

$$\begin{aligned} \frac{1}{\gamma_0}\frac{\partial \delta A_0}{\partial t} &= -(1 + i\Delta_0)\,\delta A_0 + i a_0 \nabla^2 \delta A_0, \\ \frac{1}{\gamma_1}\frac{\partial \delta A_1}{\partial t} &= -(1 + i\Delta_1)\,\delta A_1 + i a_1 \nabla^2 \delta A_1 - \overline{A_0}\,\delta A_1^*. \end{aligned} \tag{22}$$

The generic solutions of Eqs. (22) are of the form

$$(\delta A_j, \delta A_j^*) \propto e^{\lambda(k_\perp)t} e^{i\mathbf{k}_\perp \cdot \mathbf{r}_\perp}, \tag{23}$$

where $\lambda(\mathbf{k}_\perp)$ represents the growth rate of the perturbations, and $\mathbf{k}_\perp$ is the transverse component of the wavevector, which in a two-dimensional geometry obeys the relation $|\mathbf{k}_\perp|^2 = k_x^2 + k_y^2$.

Substitution of Eq. (23) in (22) and its complex conjugate, written in matrix form, allows us to evaluate the growth rates λ as the eigenvalues of the stability matrix L. This is given by

$$L = \begin{pmatrix} L_0(k_\perp) & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & L_1(k_\perp) \end{pmatrix}, \tag{24}$$

where the block matrices are defined as

$$L_0(k_\perp) = \gamma_0 \begin{pmatrix} -1 + i\Delta_0 + i a_0 k_\perp^2 & 0 \\ 0 & -1 - i\Delta_0 - i a_0 k_\perp^2 \end{pmatrix}, \tag{25}$$

and

$$L_1(k_\perp) = \gamma_1 \begin{pmatrix} -1 + i\Delta_1 + i a_1 k_\perp^2 & \dfrac{\mathcal{E}}{1 - i\Delta_0} \\ \dfrac{\mathcal{E}}{1 + i\Delta_0} & -1 - i\Delta_1 - i a_1 k_\perp^2 \end{pmatrix}. \tag{26}$$

The eigenvalues of L_0 have always a negative real part, as follows from Eq. (25), and do not predict any instability. The eigenvalues of L_1 are associated with the instability which gives rise to the subharmonic field, and are given by

$$\lambda_\pm(k_\perp) = -1 \pm \sqrt{\frac{\mathcal{E}^2}{1 + \Delta_0^2} - (\Delta_1 + a_1 k_\perp^2)^2}. \tag{27}$$

Note that only the eigenvalue with the positive sign λ_+ can take positive values and reach the instability condition $\mathrm{Re}(\lambda) > 0$.

The eigenvalue is wavenumber dependent, which means that not all the perturbations in the form of transverse plane-wave modes Eq. (23) grow at the same rate. The maximum growth rate follows from the condition $\partial\lambda_+ / \partial k = 0$. Two different cases, depending on the sign of the subharmonic detuning, must be considered:

If $\Delta_1 > 0$, which corresponds to a subharmonic frequency smaller than that of the closest cavity mode, the eigenvalue shows a maximum at

$$k_\perp = 0. \tag{28}$$

In this case, the emmited subharmonic wave travels parallel to the cavity axis, without spatial variations on the transverse plane. The solution remains homogeneous, and its amplitude is given by Eq. (19).

On the contrary, if $\Delta_1 < 0$, corresponding to a field frequency larger than that of the cavity mode, the maximum of the eigenvalue occurs at

$$k_\perp = \sqrt{-\frac{\Delta_1}{a_1}}. \tag{29}$$

The corresponding solution is of the form Eq. (23), which represents a plane wave tilted with respect to the cavity axis. This solution presents spatial variations in the transverse plane, and consequently pattern formation is expected to occur. The two kinds of instabilities are represented in Fig. 1, where the eigenvalue as a function of the wavenumber is plotted, in the cases of positive (a) and negative (b) detuning, for given values of the pump above threshold.

The value of the transverse wavenumber given by Eq. (29) can be interpreted in simple geometrical terms: $k_\perp$ corresponds to the tilt of the wave necessary to fit the longitudinal resonance condition. This is a linear effect related with diffraction, and it is represented in Fig. 2.

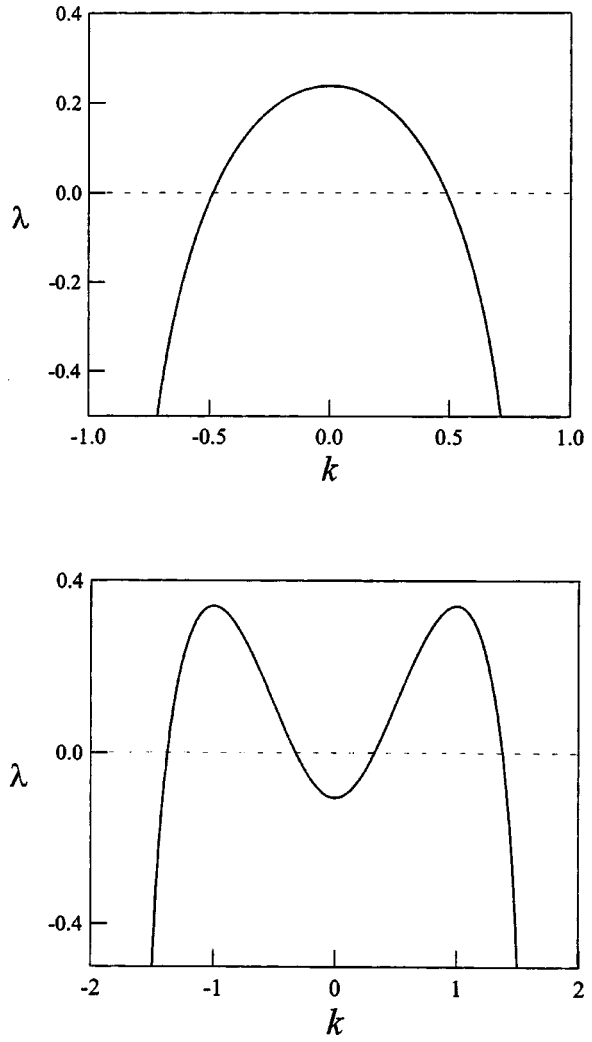


FIG. 1. Eigenvalue as a function of the perturbation wavenumber for $\Delta_1=1$, $\mathcal{E}=2.25$ (a) and $\Delta_1=-1$, $\mathcal{E}=1.35$ (b).

This can be also analytically shown by inspecting the solutions of the dynamical equations for a transverse wave in the form Eq. (23): the frequency of the cavity mode is given by $\omega=c|\mathbf{k}|=c\sqrt{k_\perp^2+k_z^2}$. Since, in the small detuning case, the relation $k_\perp \ll k_z$ holds, it can be written approximately by

$$\omega=ck_z\sqrt{1+\left(\frac{k_\perp}{k_z}\right)^2}\approx ck_z+\frac{c}{2k_z}k_\perp^2=ck_z+\Delta\omega, \qquad (30)$$

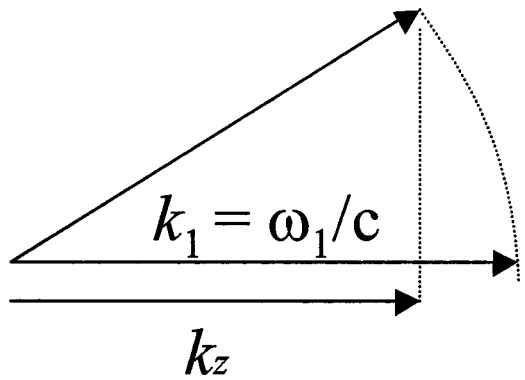


FIG. 2. Schematic representation of the longitudinal cavity resonance condition for negative detunings.

where $\Delta\omega$ is the transverse contribution to the mode frequency. This contribution arises from the diffraction term $ia\nabla^2A=i(c/2k_z)\nabla^2A$ in the dynamical equations.

Since $k_\perp$ is the modulus of the wavevector, the linear stability analysis in two dimensions predicts that a continuum of modes within a circular annulus [centered on a critical circle at $|\mathbf{k}_\perp|=k_\perp$ in (k_x,k_y) space] grows simultaneously as the pump increases above a critical value. This double infinite degeneracy of spatial modes (degenerate along a radial line from the origin and orientational degeneracy) allows, in principle, arbitrary structures in two dimensions.

The threshold (the pump value at which the instability emerges) depends also on the sign of the detuning, and is obtained from the condition Re $\lambda=0$. From Eq. (27) it follows that

$$\mathcal{E}_{th}=\sqrt{1+\Delta_0^2}\sqrt{1+(\Delta_1+a_1k^2)^2}. \qquad (31)$$

From the previous analysis, again two cases must be distinguished. For positive detunings, the mode homogeneous $k_\perp=0$ experience the maximum growth, and the threshold occurs at a pump value

$$\mathcal{E}_{th}=\sqrt{1+\Delta_0^2}\sqrt{1+\Delta_1^2}, \qquad (32)$$

which is the same found in Ref. 17. The solution above the threshold is given by Eq. (19).

For negative detunings, the instability leads to a nonhomogeneous distribution, with a characteristic scale given by the condition $a_1k_\perp^2=-\Delta_1$. The corresponding threshold for such modes is

$$\mathcal{E}_{th}=\sqrt{1+\Delta_0^2}. \qquad (33)$$

Besides the existence of a pattern forming instability, a relevant conclusion of the previous analysis is the prediction of a decrease in the threshold pump value of subharmonic generation when diffraction effects are included [compare Eqs. (32) and (33)]. This fact is especially important in acoustical systems where the nonlinearity is weak, since in this case the instability threshold, appearing for high values of the driving, can be lowered by a factor of $\sqrt{1+\Delta_1^2}$.

The predictions of the stability analysis correspond to the linear stage of the evolution, where the subharmonic field amplitude is small enough to be considered a perturbation of the trivial state. The analytical study of the further evolution would require a nonlinear stability analysis, not given here. Instead, in the next section we perform the numerical integration of Eqs. (17), where predictions of the acoustic subharmonic field in the linear and nonlinear regime are given.

V. NUMERICAL SIMULATIONS

In order to check the analytical predictions of the linear stability analysis, we integrated numerically Eqs. (17) by using the split-step technique on a spatial grid of dimensions 64×64^8. The local terms, either linear (pump, losses, and detuning) or nonlinear, are calculated in the space domain, while nonlocal terms (diffractions) are evaluated in the spatial wavevector (spectral) domain. A fast Fourier transform

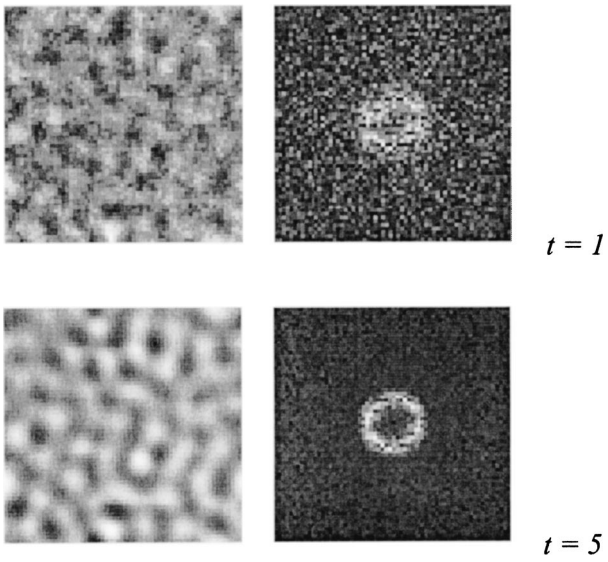


FIG. 3. The linear stage of the evolution. Shown are the pressure amplitude (left column) and the corresponding spectrum (right column) in the transverse plane. Pictures were taken at times $t=1$ (a) and $t=5$ (b). The parameters used in the integration are $\Delta_1=-2$, $\Delta_0=-1$, $\mathcal{E}=2$, $\gamma_0=\gamma_1=1$, $a_1=0.001$, and $a_0=0.0005$.

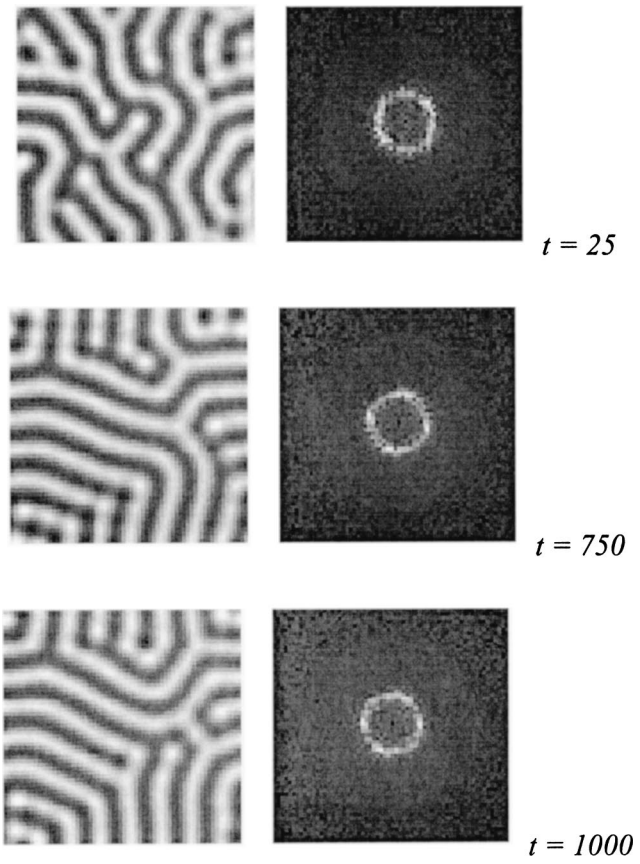


FIG. 4. Nonlinear evolution and pattern selection. The same parameters as in Fig. 3 have been used. Pictures were taken at times $t=25$ (a), $t=750$ (b), and $t=1000$ (c).

(FFT) is used to shift from spatial to spectral domains in every time step. Periodic boundary conditions are used.

As a initial contition, a noisy spatial distribution is considered, and the parameters are such that a pattern forming instability is predicted [Fig. 3(a)].

For small evolution times, the amplitude of the subharmonic remains small, corresponding to the linear stage of the evolution. As follows from the linear stability analysis, an instability ring in the far field (transverse wavenumber space) is observed [Fig. 3(b)].

For larger evolution times, the nonlinearity comes into play. In the nonlinear stage of evolution, a competition between transverse modes begins, and pattern selection is observed. In Fig. 4. it is shown a transient stage, where a labyrinthic pattern is formed. The final state, not shown in the figure, corresponds to the selection of a discrete set of transverse wavevectors, asymptotically resulting in a periodic pattern.

VI. ACOUSTICAL ESTIMATES

The theory presented in this paper has been suggested by analogies with nonlinear optical resonators, where good agreement between theory and experiment has been shown.[9] Next we estimate the required physical conditions to make the theory applicable to an acoustical nonlinear resonator.

The mean field assumption implies that the field envelopes change little during a roundtrip of the wave in the resonator. Acoustic waves, especially in the ultrasound regime, experience strong losses during propagation in the medium. It is then required that the resonator be short enough in order to avoid strong absorption.

The longitudinal size of the resonator imposes a condition on the transverse scale, in order to maintain the Fresnel number large. For example, for a pump wave of 1 MHz in water the corresponding wavelength is $\lambda \approx 1500/10^6 \approx 1.5\,\mathrm{mm}$. If the resonator length is 2 cm and the walls (transverse size of the transducer) are squared with 10 cm each side, the Fresnel number gives $F=167$, which can be considered in the large aspect ratio limit.

With this geometry, the experimental setup proposed in the first observation of parametric sound generation in a liquid filled resonator[11] seems a good candidate for the observation of the predicted phenomena. In particular, it was shown in Ref. 11 that, close to the threshold of parametric generation, only the fundamental and subharmonic frequencies were present inside the resonator, in agreement with our theory. The resonator was formed by a PZT transducer, driven at a pumping frequency ω_0, and a reflector placed at a variable distance L. The threshold was achieved when the transducer was driven at a voltage of nearly 300 V. The discrete spectrum, containing the fundamental ω and the subharmonic $\omega/2$ frequencies, existed only close to the threshold (the relative pumping level being 1 dB). In some cases, pairs of signal-idler waves (corresponding to the nondegenerate process) were excited. For larger driving voltages, an even larger number of cavity modes, at different combination frequencies, appeared due to the recurrent process of parametric amplification. The evolution of the fields at such high pump values cannot be predicted by the model proposed in this paper, which was derived under the assumption that the energy exchange occurs only between two modes. Although the spectrum is still discrete in this case, the large number of modes makes the spectral approach inconvenient for the the-

oretical study. Instead, the field approach, typically used in problems of nonlinear acoustics in nondispersive media, seems more convenient. A discussion concerning the applicability of the two approaches to the description of nonlinear wave processes in acoustics can be found in Ref. 24.

VII. CONCLUSIONS

In this paper, the problem of spontaneous emergence of patterns in an acoustical interferometer is studied from the theoretical point of view. A model for parametric sound generation in a large aspect-ratio cavity is derived, taking into account the effects of diffraction. It is shown that the subharmonic field can be excited, when a threshold pump value is reached, characterized by a nonuniform distribution in the transverse plane or pattern. This occurs when the field frequency is tuned below the frequency of the closest resonator mode (negative detuning). On the contrary, an on-axis field, with homogeneous distribution, is emmited. Traditionally, patterns arise from the imposition of external constraints (waveguiding). The field then oscillates in modes of the resonator. The spontaneous emergence of patterns considered in this article presupposes no external transverse mode selection mechanism, but instead allows the system to choose a pattern through the nonlinear interaction of a tipically infinite set of degenerate modes. The pattern formation process described here is related with the competing effects of nonlinearity and diffraction, and presents many analogies with similar systems studied in nonlinear optics, such as the two-level laser or the optical parametric oscillator, to which the model derived in this paper is isomorphous.

ACKNOWLEDGMENTS

The author thanks Dr. Y. N. Makov, V. E. Gusev, V. Espinosa, J. Ramis, and J. Alba for interesting discussions on the subject. The work was financially supported by the CICYT of the Spanish Government, under Project Nos. PB98-0635-C03-02 and BFM2002-04369-C04-04.

APPENDIX: DERIVATION OF THE EVOLUTION EQUATIONS EQS. (12)

In this appendix we derive the model for parametric sound generation given in Eqs. (17). The approach is based on the method proposed in Ref. 25 to describe the dynamics of laser fields in Fabry–Perot resonators.

The first assumption is that the longitudinal variations of the fields are mainly accounted for by the plane wave, in which case the envelope amplitudes F_j and B_j can be considered as slowly varying functions in z and t. This leads to the inequalities

$$\left|\frac{\partial^2 F_j}{\partial z^2}\right| \ll \left|k_j \frac{\partial F_j}{\partial z}\right|, \quad \left|\frac{\partial^2 F_j}{\partial t^2}\right| \ll \left|\omega_j \frac{\partial F_j}{\partial t}\right|, \tag{A1}$$

which allow us to neglect the second-order derivatives in z and t.

In the same way, applying the second relation of Eq. (A1) to the dissipation term, it reduces to

$$\frac{\delta_j}{c_j^4}\frac{\partial^3 p_j}{\partial t^3} \approx i\frac{\delta_j}{c_j^4}\omega_j^3 p_j. \tag{A2}$$

The condition given by Eq. (A1) leads to the well-known parabolic or eikonal approximation, in which the d'Alembertian operator acting on the pressure on the left-hand side of Eq. (11) can be approximated by

$$\nabla^2 - \frac{1}{c_j^2}\frac{\partial^2}{\partial t^2} \approx 2ik_j\left(\frac{\partial}{\partial z} + \frac{1}{c_j}\frac{\partial}{\partial t}\right) + \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right). \tag{A3}$$

With Eqs. (A2) and (A3) the wave evolution can be written as

$$\begin{aligned} &e^{ik_j z}\left(c_j\frac{\partial F_j}{\partial z} + \frac{\partial F_j}{\partial t} + \frac{c_j}{2ik_j}\nabla_\perp^2 F_j + \kappa_j F_j\right) \\ &+ e^{-ik_j z}\left(-c_j\frac{\partial B_j}{\partial z} + \frac{\partial B_j}{\partial t} + \frac{c_j}{2ik_j}\nabla_\perp^2 B_j + \kappa_j B_j\right) \\ &= -ib_j[p^2]_j^-, \end{aligned} \tag{A4}$$

where $\kappa_j = \delta_j k_j^2/2$ is a parameter that accounts for dissipation, $b_j = \beta k_j/2\rho_0 c_j$ is a nonlinearity parameter, and where $[p^2]_j^-$ contains the terms in p^2 that oscillate with frequency $-\omega_j$, and takes into account slow and fast spatial variations.

The nonlinear terms at the frequencies ω_0 and ω_1 are, respectively,

$$\begin{aligned} [p^2]_0^- &= \tfrac{1}{4}B_1^2 e^{-ik_0 z} + \tfrac{1}{2}B_1 F_1 + \tfrac{1}{4}F_1^2 e^{ik_0 z}, \\ [p^2]_1^- &= \tfrac{1}{2}B_1^* B_0 e^{-ik_1 z} + \tfrac{1}{2}F_1^* B_0 e^{-3ik_1 z} + \tfrac{1}{2}B_1^* F_0 e^{3ik_1 z} \\ &\quad + \tfrac{1}{2}F_1^* F_0 e^{ik_1 z}. \end{aligned} \tag{A5}$$

In order to eliminate the explicit dependence of the exponential factors in Eqs. (A5), a projection on two longitudinal modes is performed, multiplying by $(1/2\pi)\exp(\pm ik_j z)$, and integrating over a full wavelength. This leads to a system of equations where the fields F_j and B_j are decoupled in the linear part (lhs), i.e., we get

$$\begin{aligned} &c_j\frac{\partial F_j}{\partial z} + \frac{\partial F_j}{\partial t} + \frac{c_j}{2ik_j}\nabla_\perp^2 F_j + \kappa_j F_j \\ &= -i\frac{b_j}{k_j}\frac{1}{2\pi}\int_{-\pi}^{\pi}[p^2]_j^- e^{-ik_j z}\, d(k_j z) \end{aligned} \tag{A6}$$

for the forward waves, and

$$\begin{aligned} &-c_j\frac{\partial B_j}{\partial z} + \frac{\partial B_j}{\partial t} + \frac{c_j}{2ik_j}\nabla_\perp^2 B_j + \kappa_j B_j \\ &= -i\frac{b_j}{k_j}\frac{1}{2\pi}\int_{-\pi}^{\pi}[p^2]_j^- e^{ik_j z}\, d(k_j z) \end{aligned} \tag{A7}$$

for the backward waves.

Substituting the nonlinear terms (A5), only one of the contributions survives, leading to a couple of equations for each frequency,

$$c_0\frac{\partial F_0}{\partial z} + \frac{\partial F_0}{\partial t} + \frac{c_0}{2ik_0}\nabla_\perp^2 F_0 + \kappa_0 F_0 = -i\frac{b_0}{4}F_1^2, \tag{A8}$$

$$-c_0\frac{\partial B_0}{\partial z}+\frac{\partial B_0}{\partial t}+\frac{c_0}{2ik_0}\nabla_\perp^2 B_0+\kappa_0 B_0=-i\frac{b_0}{4}B_1^2,$$

for the fundamental wave, and

$$\begin{aligned} c_1\frac{\partial F_1}{\partial z}+\frac{\partial F_1}{\partial t}+\frac{c_1}{2ik_1}\nabla_\perp^2 F_1+\kappa_1 F_1&=-i\frac{b_1}{2}F_1^* F_0,\\ -c_1\frac{\partial B_1}{\partial z}+\frac{\partial B_1}{\partial t}+\frac{c_1}{2ik_1}\nabla_\perp^2 B_1+\kappa_1 B_1&=-i\frac{b_1}{2}B_1^* B_0, \end{aligned} \tag{A9}$$

for the subharmonic.

Besides the existence of the counter-propagating waves, the cavity also imposes a condition which relates the amplitudes of forward and backward in given points of the resonator. The fields obey the following boundary conditions in the cavity: in $z=0$,

$$\begin{aligned} F_0(x,y,0,t)&=\sqrt{\mathcal{R}_0}B_0(x,y,0,t)+\sqrt{\mathcal{T}_0}Y,\\ F_1(x,y,0,t)&=\sqrt{\mathcal{R}_1}B_1(x,y,0,t), \end{aligned} \tag{A10}$$

and, in $z=L$,

$$\begin{aligned} B_0(x,y,L,t)&=\sqrt{\mathcal{R}_0}e^{-i\theta_0}F_0(x,y,L,t),\\ B_1(x,y,L,t)&=\sqrt{\mathcal{R}_1}e^{-i\theta_1}F_1(x,y,L,t), \end{aligned} \tag{A11}$$

where L is the length of the cavity, $\mathcal{T}_0$ is the transmitivity of the boundary to the input wave (pump), with amplitude given by Y, and $\mathcal{R}_j$ is the reflectivity of the field p_j at the boundary. Finally, θ_j are the detunings (frequency mismatch with respect to the cavity), given by

$$\theta_0=2\pi m-2k_0L=\frac{\omega_0^c-\omega_0}{c_0/2L}, \tag{A12}$$

$$\theta_1=2\pi n-2k_1L=\frac{\omega_1^c-\omega_1}{c_1/2L}, \tag{A13}$$

ω_j^c being the frequency of the cavity (eigenmode) nearest to the field frequency ω_j.

The field evolution is described by Eqs. (A8) and (A9) together with Eqs. (A10) and (A11). However, the description can be greatly simplified after the introduction of the following changes:[26]

$$\begin{aligned} \tilde{F}_0&=e^{(1/2L)(z-L)(\ln\mathcal{R}_0-i\theta_0)}F_0+\frac{\sqrt{\mathcal{T}_0}}{\sqrt{\mathcal{R}_0}}e^{i(\theta_0/2)}\frac{1}{2L}(z-L)Y,\\ \tilde{B}_0&=e^{-(z/2L)(\ln\mathcal{R}_0-i\theta_0)}B_0e^{i(\theta_0/2)}\\ &\quad-\frac{\sqrt{\mathcal{T}_0}}{\sqrt{\mathcal{R}_0}}e^{i(\theta_0/2)}\frac{1}{2L}(z-L)Y,\\ \tilde{F}_1&=e^{(1/2L)(z-L)(\ln\mathcal{R}_1-i\theta_1)}F_1,\\ \tilde{B}_1&=e^{-(z/2L)(\ln\mathcal{R}_1-i\theta_1)}B_1e^{i(\theta_1/2)}, \end{aligned} \tag{A14}$$

The great advantage of the latter changes is that the boundary conditions for these new fields are

$$\begin{aligned} \tilde{F}_0(0,t)&=\tilde{B}_0(0,t),\\ \tilde{F}_1(L,t)&=\tilde{B}_1(L,t), \end{aligned} \tag{A15}$$

which correspond to those of an ideal cavity with perfectly reflecting boundaries. This fact will be used later for determining the unknow longitudinal distribution.

Substituting the new amplitudes in the evolutions equations we find the new system

$$\begin{aligned} \frac{\partial\tilde{F}_0}{\partial t}+c_0\frac{\partial\tilde{F}_0}{\partial z}&=-c_0\frac{|\ln\mathcal{R}_0|+i\theta_0}{2L}\tilde{F}_0\\ &\quad+c_0\frac{|\ln\mathcal{R}_0|+i\theta_0}{4L^2}(z-L)\frac{\sqrt{\mathcal{T}_0}}{\mathcal{R}_0}Ye^{i(\theta_0/2)}\\ &\quad+\frac{c_0}{2L}\frac{\sqrt{\mathcal{T}_0}}{\sqrt{\mathcal{R}_0}}Ye^{i(\theta_0/2)}-\frac{c_0}{2ik_0}\nabla^2\tilde{F}_0\\ &\quad-\kappa_0\tilde{F}_0-i\frac{b_0}{4}\mathcal{D}(z)\tilde{F}_1^2, \end{aligned} \tag{A16}$$

where we have defined the exponential term

$$\mathcal{D}(z)=e^{(1/2L)(z-L)(\ln\mathcal{R}_0-i\theta_0)}e^{(-2/2L)(z-L)(\ln\mathcal{R}_1-i\theta_1)}. \tag{A17}$$

Consider now two additional conditions: first, that the reflectivity at the boundaries is high (weak losses), which is expressed as $\mathcal{R}_0\to 1$, and consequently $\mathcal{T}_0\to 0$, and, second, that the field frequencies are close to a resonant frequency of the cavity, so that $\theta_0\to 0$. Under these conditions, also known as the mean-field limit in nonlinear optics, the exponential factor $\mathcal{D}(z)$ approaches unity, and the second term at the lhs of Eq. (A16) can be neglected. Furthermore, in this limit the field amplitudes $\tilde{F}_j$ and $\hat{B}_j$ approach to their real original, F_j and B_j. The equation (A16) takes then a simplified form, which can be more conveniently written as

$$\frac{\partial F_0}{\partial t}+c_0\frac{\partial F_0}{\partial z}=E-\gamma_0(1+i\Delta_0)F_0+ia_0\nabla^2F_0-i\frac{b_0}{4}F_1^2, \tag{A18}$$

where we have defined the new pump E, detuning Δ_0, losses γ_0, and diffraction a_0 parameters as

$$E=\frac{c_0}{2L}\frac{\sqrt{\mathcal{T}_0}}{\sqrt{\mathcal{R}_0}}Ye^{i(\theta_0/2)}, \tag{A19}$$

$$\Delta_j=\frac{\theta_j}{|\ln\mathcal{R}_j|+(2L/c_j)\kappa_j}, \tag{A20}$$

$$\gamma_j=\frac{|\ln\mathcal{R}_j|c_j}{2L}+\kappa_j, \tag{A21}$$

$$a_j=\frac{c_j}{2k_j}, \tag{A22}$$

where $j=0,1$. Defined in this way, the detuning is a quantity of the order of 1. Also, γ_0 represents a measure of the linewidth of the cavity modes.

We have performed the derivation for the forward component of the fundamental wave. A similar procedure leads to the evolution equations for the other components,

$$\frac{\partial B_0}{\partial t}+c_0\frac{\partial B_0}{\partial z}=E-\gamma_0(1+i\Delta_0)B_0+ia_0\nabla^2B_0-i\frac{b_0}{4}B_1^2, \tag{A23}$$

$$\frac{\partial F_1}{\partial t}+c_1\frac{\partial F_1}{\partial z}=-\gamma_1(1+i\Delta_1)F_1+ia_1\nabla^2F_1 -i\frac{b_1}{2}F_1^*F_0, \tag{A24}$$

$$\frac{\partial B_1}{\partial t}+c_1\frac{\partial B_1}{\partial z}=-\gamma_1(1+i\Delta_1)B_1+ia_1\nabla^2B_1 -i\frac{b_1}{2}B_1^*B_0. \tag{A25}$$

The equations still keep an explicit z dependence. Consider now the new boundary conditions, given by Eq. (A15). These conditions differ from that of the original fields in that they represent an ideal (lossless) cavity, which allows us to express the field inside the cavity by means of the Fourier expansions

$$\hat{F}_j(\mathbf{r},t)=\sum_{n=-\infty}^{\infty}p_j^{(n)}(x,y,t)\exp\left(i\frac{n\pi z}{L}\right),$$
$$\hat{B}_j(\mathbf{r},t)=\sum_{n=-\infty}^{\infty}p_j^{(n)}(x,y,t)\exp\left(-i\frac{n\pi z}{L}\right). \tag{A26}$$

From (9), it follows that the total field with frequency ω_j can be also written as

$$P_j(\mathbf{r},t)=\sum_{n=-\infty}^{\infty}2p_j^{(n)}(x,y,t)\cos\left[\left(k_j+\frac{n\pi}{L}\right)z\right]e^{-i\omega_jt} +\text{c.c.} \tag{A27}$$

If we finally assume that the intermode spacing $c/2L$ is much larger than the detuning $\omega_c-\omega$, we can consider that only the $n=0$ mode can be excited, and F_0 and B_0 become spatially uniform along the longitudinal axis of the cavity. In this case, the field can be described by (8), which is the usual description for waves in acoustic resonators.

This last assumption makes the dynamical evolution to be independent on z, and Eq. (A18) take the final form

$$\frac{\partial p_0}{\partial t}=E-\gamma_0(1+i\Delta_0)p_0+ia_0\nabla^2p_0-i\frac{b_0}{4}p_1^2,$$
$$\frac{\partial p_1}{\partial t}=-\gamma_1(1+i\Delta_1)p_1+ia_1\nabla^2p_1-i\frac{b_1}{2}p_1^*p_0, \tag{A28}$$

which are the equations (12) given in the text.

[1] M. C. Cross and P. C. Hohenberg, "Pattern formation outside of equilibrium," Rev. Mod. Phys. **65**, 851–1112 (1993).

[2] I. Prigogine and G. Nicolis, "On symmetry-breaking instabilities in dissipative systems," J. Chem. Phys. **46**, 3542–3550 (1967).

[3] L. A. Ostrovsky and I. A. Soustova, "Theory of parametric sound generators," Sov. Phys. Acoust. **22**, 416–419 (1976).

[4] J. W. Miles, "Nonlinear Faraday resonance," J. Fluid Mech. **146**, 285–302 (1984).

[5] F. Melo, P. Umbanhowar, and H. L. Swinney, "Transition to parametric wave patterns in a vertically oscillated granular layer," Phys. Rev. Lett. **72**, 172–175 (1994).

[6] V. L'vov, *Wave Turbulence Under Parametric Excitation* (Springer-Verlag, Berlin, 1994).

[7] G.-L. Oppo, M. Brambilla, and L. A. Lugiato, "Formation and evolution of roll patterns in optical parametric oscillators," Phys. Rev. A **49**, 2028–2032 (1994).

[8] G. J. de Valcárcel, K. Staliunas, E. Roldán, and V. J. Sánchez-Morcillo, "Transverse patterns in degenerate optical parametric oscillators and degenerate four-wave mixing," Phys. Rev. A **54**, 1609–1624 (1996).

[9] F. T. Arecchi, S. Boccaletti, and P.-L. Ramazza, "Pattern formation and competition in nonlinear optics," Phys. Rep. **318**, 1–83 (1999).

[10] F. V. Bunkin, Y. A. Kravtsov, and G. A. Lyskhov, "Acoustic analogues of nonlinear-optics phenomcna," Sov. Phys. Usp. **29**, 607–619 (1986).

[11] A. Korpel and R. Adler, "Parametric phenomena observed on ultrasonic waves in water," Appl. Phys. Lett. **7**, 106–108 (1965).

[12] L. Adler and M. A. Breazeale, "Generation of fractional harmonics in a resonant untrasonic wave system," J. Acoust. Soc. Am. **48**, 1077–1083 (1970).

[13] N. Yen, "Experimental investigation of subharmonic generation in an acoustic interferometer," J. Acoust. Soc. Am. **57**, 1357–1362 (1975).

[14] D. Rinberg, V. Cherepanov, and V. Steinberg, "Parametric generation of second sound by first sound in superfluid helium," Phys. Rev. Lett. **76**, 2105–2108 (1996).

[15] D. Rinberg and V. Steinberg, "Parametric generation of second sound in superfluid helium: linear stability and nonlinear dynamics," Phys. Rev. B **64**, 054506 (2001).

[16] M. F. Hamilton and J. A. TenCate, "Sum and difference frequency generation due to noncollinear wave interaction in a rectangular duct," J. Acoust. Soc. Am. **81**, 1703–1712 (1987).

[17] L. A. Ostrovsky, I. A. Soustova, and A. M. Sutin, "Nonlinear and parametric phenomena in dispersive acoustic systems," Acustica **39**, 298–306 (1978).

[18] M. F. Hamilton and D. Blackstock, *Nonlinear Acoustics* (Academic, New York, 1997).

[19] L. K. Zarembo and O. Y. Serdoboloskaya, "To the problem of parametric amplification and parametric generation of acoustic waves," Akust. Zh. **20**, 726–732 (1974).

[20] I. Akhatov, U. Parlizt, and W. Lauterborn, "Pattern formation in acoustic cavitation," J. Acoust. Soc. Am. **96**, 3627–3635 (1994).

[21] I. Akhatov, U. Parlizt, and W. Lauterborn, "Towards a theory of self-organization phenomena in bubble-liquid mixtures," Phys. Rev. E **54**, 4990–5003 (1996).

[22] P. J. Westervelt, "Parametric acoustic array," J. Acoust. Soc. Am. **35**, 535–537 (1963).

[23] O. A. Druzhinin, L. A. Ostrovsky, and A. Prosperetti, "Low frequency acoustic wave generation in a resonant bubble layer," J. Acoust. Soc. Am. **100**, 3570–3580 (1996).

[24] L. A. Ostrovsky, A. M. Sutin, I. A. Soustova, A. I. Matveyev, and A. I. Potapov, "Nonlinear, low-frequency sound generation in a bubble layer: Theory and laboratory experiment," J. Acoust. Soc. Am. **104**, 722–726 (1998).

[25] O. V. Rudenko, "Nonlinear sawtooth-shaped waves," Sov. Phys. Usp. **38**, 965–990 (1995).

[26] L. A. Lugiato and L. M. Narducci, "Nonlinear dynamics in a Fabry-Perot resonator," Z. Phys. B: Condens. Matter **71**, 129–138 (1988).

Spherical wave propagation through inhomogeneous, anisotropic turbulence: Log-amplitude and phase correlations

Vladimir E. Ostashev
NOAA/Environmental Technology Laboratory, 325 Broadway, Boulder, Colorado 80305
and Department of Physics, New Mexico State University, Las Cruces, New Mexico 88003

D. Keith Wilson
U.S. Army Research Laboratory, 2800 Powder Mill Road, Adelphi, Maryland 20783-1197

George H. Goedecke
Department of Physics, New Mexico State University, Las Cruces, New Mexico 88003

(Received 12 March 2002; revised 8 September 2003; accepted 23 September 2003)

Inhomogeneity and anisotropy are intrinsic characteristics of daytime and nighttime turbulence in the atmospheric boundary layer. In the present paper, line-of-sight sound propagation through inhomogeneous, anisotropic turbulence with temperature and velocity fluctuations is considered. Starting from a parabolic equation and using the Markov approximation, formulas are derived for the correlation functions and variances of log-amplitude and phase fluctuations of a spherical sound wave. These statistical moments of a sound field are important for many practical applications in atmospheric acoustics. The derived formulas for the correlation functions and variances generalize those already known in the literature for two limiting cases: (a) homogeneous, isotropic turbulence, and (b) inhomogeneous, anisotropic turbulence with temperature fluctuations only. Furthermore, the formulas differ from those for the case of plane wave propagation. Using the derived formulas and Mann's spectral tensor of velocity fluctuations for shear-driven turbulence, the correlation functions and variances of log-amplitude and phase fluctuations are studied numerically. The results obtained clearly show that turbulence inhomogeneity and anisotropy significantly affect sound propagation in the atmosphere. [DOI: 10.1121/1.1628680]

PACS numbers: 43.28.Gq, 43.28.Lv [LCS] Pages: 120–130

I. INTRODUCTION

Studies of sound propagation in a turbulent atmosphere are important in many practical problems, such as noise pollution near airports and highways, acoustic remote sensing, performance of modern acoustical systems for source detection, ranging and recognition, and sonic boom propagation. Theories of sound propagation through homogeneous and isotropic turbulence with temperature and velocity fluctuations have undergone significant development in the past several years and have been summarized in Ref. 1. These theories generalize classical theories of wave propagation in random media with temperature fluctuations only.[2–6]

Turbulence can be modeled as homogeneous and isotropic if statistical moments of a sound field are mainly affected by small-scale eddies in the inertial subrange. Large-scale (energy-containing subrange) eddies in daytime and nighttime atmospheric boundary layers (ABL) can also affect statistical moments of a sound field. Intrinsic characteristics of the large-scale eddies are anisotropy and inhomogeneity.[7] Anisotropy is caused by stretching and other distortions of eddies due to shear and buoyancy forces. Inhomogeneity of atmospheric turbulence here refers to dependence of turbulence statistics on spatial position, in particular the height above the ground. References 8–11 have shown that turbulence anisotropy and inhomogeneity can significantly affect the statistical properties of a sound field propagating in the atmosphere. Wave propagation through inhomogeneous and anisotropic turbulence with temperature fluctuations was also studied in classical theories (e.g., Ref. 2).

A previous paper[11] considered line-of-sight propagation of a plane sound wave through inhomogeneous, anisotropic turbulence with temperature and velocity fluctuations. Formulas for the correlation functions and variances of log-amplitude and phase fluctuations of the plane wave were derived in that reference. These formulas were then used for numerical studies of the effects of turbulence inhomogeneity and anisotropy on log-amplitude and phase fluctuations.

The goal of the present paper is to generalize the results obtained in Ref. 11 to the case of spherical wave propagation.[12] The line-of-sight geometry we use for this problem is shown in Fig. 1. A point source is located at the origin of the Cartesian coordinate system x,y,z. A sound field due to this source is monitored by two receivers from which the correlation functions of the log-amplitude and phase fluctuations are determined. (Only one receiver is needed to calculate the variances of the log-amplitude and phase fluctuations.) The geometrical center of the two receivers is located close to the x axis. Sound waves reflected from the ground are ignored. (Reflected waves can, for example, be eliminated by placing receivers on the ground while a source is elevated, or by using impulse source.)

Note that the statistical moments of spherical and plane sound waves propagating through the same turbulence are different. These two limiting cases are useful for different applications. In situations where both the source and receiver

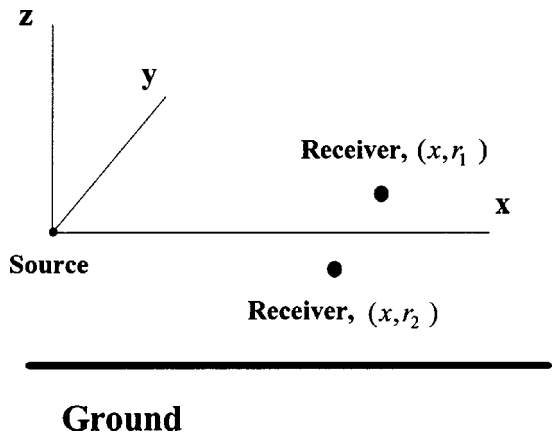


FIG. 1. Geometry of line-of-sight sound propagation from source to two receivers.

are located within the ABL (which is about 1 km in height), the spherical wave propagation case is most nearly valid. However, if the sound source is located much higher than 1 km and the receiver is near the ground, one can assume that a plane sound wave is incident on the turbulent atmosphere within the ABL, because temperature and velocity fluctuations above the ABL are usually very small. Therefore, in the literature, it is common to consider separately the statistical moments of plane and spherical waves, e.g., Refs. 1, 4–6, 13, and 14.

Also note that formulas for the variances and correlation functions of log-amplitude and phase fluctuations of a sound wave propagating in a turbulent atmosphere derived in Ref. 11 and in the present paper are important for many practical applications in atmospheric acoustics. For example, these formulas can be used for modeling the performance of modern beamforming arrays for source localization in the atmosphere.[15,16]

In this paper, the log-amplitude and phase fluctuations of a spherical sound wave propagating through inhomogeneous, anisotropic turbulence are expressed in terms of temperature and velocity fluctuations using the Rytov method (Sec. II). Formulas for the correlation functions and variances of the log-amplitude and phase fluctuations of a spherical sound wave are then derived using the Markov approximation (Sec. III). A velocity spectrum of inhomogeneous, anisotropic, shear-driven turbulence is presented in Sec. IV. Using this spectrum, the effects of turbulence inhomogeneity and anisotropy on the correlation functions and variances of log-amplitude and phase fluctuations are numerically studied in Sec. V. The results obtained are compared with those for plane wave propagation. Conclusions are presented in Sec. VI.

II. LOG-AMPLITUDE AND PHASE FLUCTUATIONS

In this section, formulas for the log-amplitude and phase fluctuations of a spherical sound wave propagating through inhomogeneous, anisotropic turbulence are derived.

A. Parabolic equation

Let $p(x,\mathbf{r})$ be the sound pressure of a spherical sound wave, where $\mathbf{r}=(y,z)$ is the vector of transverse coordinates with respect to the x axis. It is worthwhile to express p in terms of the complex amplitude A of this wave:

$$p(x,\mathbf{r})=e^{ikx}A(x,\mathbf{r}), \tag{1}$$

where k is the sound wave number. We will assume that the complex amplitude A satisfies the following narrow-angle parabolic equation:[1]

$$2ik\frac{\partial A(x,\mathbf{r})}{\partial x}+\Delta_{\perp}A(x,\mathbf{r})+k^2\varepsilon_{\mathrm{mov}}(x,\mathbf{r})A(x,\mathbf{r})=0. \tag{2}$$

In this equation, $\Delta_{\perp}=\partial^2/\partial y^2+\partial^2/\partial z^2$, and the random field $\varepsilon_{\mathrm{mov}}(x,\mathbf{r})=-\tilde{T}(x,\mathbf{r})/T_0-2v_x(x,\mathbf{r})/c_0$ is twice the fluctuation in acoustic refractive index in a turbulent atmosphere. Here, $\tilde{T}$ is the temperature fluctuation, v_x is the fluctuation in the x-component of the wind velocity vector $\mathbf{v}$, and T_0 and c_0 are the mean values of the temperature and adiabatic sound speed.

When considering the log-amplitude and phase fluctuations, it is convenient to express the complex amplitude A as

$$A(x,\mathbf{r})=e^{\Psi(x,\mathbf{r})}A_0(x,\mathbf{r}). \tag{3}$$

Here, Ψ is the fluctuation of the complex phase due to temperature and velocity fluctuations in the atmosphere, and A_0 is the complex amplitude of a sound field in the absence of these fluctuations. The complex phase can be written as a sum: $\Psi=\chi+i\phi$, where $\chi(x,\mathbf{r})$ and $\phi(x,\mathbf{r})$ are the log-amplitude and phase fluctuations. These can be expressed in terms of the complex phase:

$$\chi(x,\mathbf{r})=\tfrac{1}{2}[\Psi(x,\mathbf{r})+\Psi^*(x,\mathbf{r})], \tag{4}$$

$$\phi(x,\mathbf{r})=\frac{1}{2i}[\Psi(x,\mathbf{r})-\Psi^*(x,\mathbf{r})]. \tag{5}$$

B. Complex phase

Substituting Eq. (3) for the complex amplitude A into the parabolic equation (2) and using the Rytov approximation, one obtains a formula for the complex phase Ψ in terms of the random field $\varepsilon_{\mathrm{mov}}$ [see Eq. (20) from Ref. 11]:

$$\Psi(x,\mathbf{r})=\frac{k^2}{4\pi}\int_0^x\frac{dx'}{x-x'}\int d\mathbf{r}'\varepsilon_{\mathrm{mov}}(x',\mathbf{r}')\times\exp\left(\frac{ik(\mathbf{r}-\mathbf{r}')^2}{2(x-x')}\right)\frac{A_0(x',\mathbf{r}')}{A_0(x,\mathbf{r})}. \tag{6}$$

Here, x' and $\mathbf{r}'=(y',z')$ are the integration variables along the x axis and in the plane perpendicular to it. The Rytov approximation is valid if the variance of log-amplitude fluctuations is less than one. For the case of plane wave propagation considered in Ref. 11, the ratio of A_0 terms in the integrand in Eq. (6) is 1. For spherical wave propagation, this ratio is calculated below.

In the absence of temperature and velocity fluctuations, the sound pressure p_0 due to a point source located at the origin of the coordinate system (see Fig. 1) is given by

$$p_0(x,\mathbf{r})=p^0\frac{\exp(ik\sqrt{x^2+r^2})}{\sqrt{x^2+r^2}}, \tag{7}$$

where p^0 is a factor proportional to the sound pressure amplitude. In the approximation of the parabolic equation, $r \ll x$. Therefore, in this approximation, the sound pressure p_0 can be written as

$$p_0(x,\mathbf{r}) = p^0 \frac{\exp(ikx + ikr^2/(2x))}{x}. \tag{8}$$

Substituting this formula into Eq. (1), we obtain the following expression for the complex amplitude of a spherical wave propagating in the atmosphere without temperature and velocity fluctuations: $A_0(x,\mathbf{r}) = (p^0/x)\exp[ikr^2/(2x)]$. The use of this expression in Eq. (6) results in a formula for the complex phase $\boldsymbol{\Psi}$ of a spherical sound wave propagating in a turbulent atmosphere with temperature and velocity fluctuations

$$\boldsymbol{\Psi}(x,\mathbf{r}) = \frac{xk^2}{4\pi} e^{-ikr^2/2x} \int_0^x \frac{dx'}{x'(x-x')} \int d\mathbf{r}' \varepsilon_{\mathrm{mov}}(x',\mathbf{r}') \times \exp\left[\frac{ik(\mathbf{r}-\mathbf{r}')^2}{2(x-x')} + \frac{ik(\mathbf{r}')^2}{2x'}\right]. \tag{9}$$

For the case of a turbulent atmosphere with temperature fluctuations only, this equation is essentially the same as Eq. (2.74) from Ref. 17.

Substitution of Eq. (9) into Eqs. (4) and (5) allows one to obtain formulas for the log-amplitude and phase fluctuations, $\chi(x,\mathbf{r})$ and $\phi(x,\mathbf{r})$, in terms of the random field $\varepsilon_{\mathrm{mov}}(x,\mathbf{r})$. These formulas are not presented here since it is more convenient to deal with their spectral analogs when deriving equations for the correlation functions and variances of χ and ϕ. The next two subsections present derivations of formulas for the complex phase and the log-amplitude and phase fluctuations in a spectral domain.

C. Spectral representation of the complex phase

The random fields $\boldsymbol{\Psi}(x,\mathbf{r})$ and $\varepsilon_{\mathrm{mov}}(x,\mathbf{r})$ are expressed as the two-dimensional Fourier integrals:

$$\boldsymbol{\Psi}(x,\mathbf{r}) = \int d\mathbf{K}_\perp e^{i\mathbf{K}_\perp\cdot\mathbf{r}} \hat{\boldsymbol{\Psi}}(x,\mathbf{K}_\perp), \tag{10}$$

$$\varepsilon_{\mathrm{mov}}(x,\mathbf{r}) = \int d\mathbf{K}_\perp e^{i\mathbf{K}_\perp\cdot\mathbf{r}} \hat{\varepsilon}_{\mathrm{mov}}(x,\mathbf{K}_\perp). \tag{11}$$

Here, $\hat{\boldsymbol{\Psi}}(x,\mathbf{K}_\perp)$, and $\hat{\varepsilon}_{\mathrm{mov}}(x,\mathbf{K}_\perp)$ are the two-dimensional spatial spectral densities of the corresponding random fields.

In Eq. (9), $\boldsymbol{\Psi}$ and $\varepsilon_{\mathrm{mov}}$ are replaced by their values given by Eqs. (10) and (11). After some algebra, we obtain a formula for the two-dimensional spectral density of the complex phase in terms of $\hat{\varepsilon}_{\mathrm{mov}}$

$$\hat{\boldsymbol{\Psi}}(x,\mathbf{K}_\perp) = \frac{ik}{2} \int_0^x dx' \hat{\varepsilon}_{\mathrm{mov}}\left(x', \frac{x\mathbf{K}_\perp}{x'}\right) \frac{x^2}{(x')^2} \times \exp\left[-\frac{iK_\perp^2 x(x-x')}{2kx'}\right]. \tag{12}$$

Since $\varepsilon_{\mathrm{mov}}$ is a real random field, the following equality holds:

$$\hat{\varepsilon}^*_{\mathrm{mov}}(x,\mathbf{K}_\perp) = \hat{\varepsilon}_{\mathrm{mov}}(x,-\mathbf{K}_\perp). \tag{13}$$

This equality can be proved by complex conjugation of both sides of Eq. (11) and equating the right-hand side of the resulting expression to the right-hand side of Eq. (11).

Now conjugating both sides of Eq. (12) and using Eq. (13), we obtain another formula for the spectral density of the complex phase:

$$\hat{\boldsymbol{\Psi}}^*(x,-\mathbf{K}_\perp) = -\frac{ik}{2} \int_0^x dx' \hat{\varepsilon}_{\mathrm{mov}}\left(x', \frac{x\mathbf{K}_\perp}{x'}\right) \frac{x^2}{(x')^2} \times \exp\left[\frac{iK_\perp^2 x(x-x')}{2kx'}\right]. \tag{14}$$

In the next subsection, Eqs. (12) and (14) are used to derive formulas for log-amplitude and phase fluctuations in a spectral domain.

D. Spectral representations of log-amplitude and phase fluctuations

Let us express $\chi(x,\mathbf{r})$ and $\phi(x,\mathbf{r})$ as the two-dimensional Fourier integrals:

$$\chi(x,\mathbf{r}) = \int d\mathbf{K}_\perp e^{i\mathbf{K}_\perp\cdot\mathbf{r}} \hat{\chi}(x,\mathbf{K}_\perp), \tag{15}$$

$$\phi(x,\mathbf{r}) = \int d\mathbf{K}_\perp e^{i\mathbf{K}_\perp\cdot\mathbf{r}} \hat{\phi}(x,\mathbf{K}_\perp). \tag{16}$$

In these equations, $\hat{\chi}(x,\mathbf{K}\perp)$ and $\hat{\phi}(x,\mathbf{K}_\perp)$ are the two-dimensional spectral densities of log-amplitude and phase fluctuations.

Replacing $\boldsymbol{\Psi}$ and χ in Eq. (4) by their values from the right-hand sides of Eqs. (10) and (15), $\hat{\chi}$ can be expressed in terms of $\hat{\Psi}$:

$$\hat{\chi}(x,\mathbf{K}_\perp) = \tfrac{1}{2}[\hat{\boldsymbol{\Psi}}(x,\mathbf{K}_\perp) + \hat{\boldsymbol{\Psi}}^*(x,-\mathbf{K}_\perp)]. \tag{17}$$

A formula for $\hat{\phi}$ is derived similarly:

$$\hat{\phi}(x,\mathbf{K}_\perp) = \frac{1}{2i}[\hat{\boldsymbol{\Psi}}(x,\mathbf{K}_\perp) - \hat{\boldsymbol{\Psi}}^*(x,-\mathbf{K}_\perp)]. \tag{18}$$

Substituting the values of $\hat{\Psi}$ and $\hat{\Psi}^*$ given by Eqs. (12) and (14) into Eqs. (17) and (18), we obtain the desired formulas for the spectral densities of log-amplitude and phase fluctuations:

$$\hat{\chi}(x,\mathbf{K}_\perp) = \frac{k}{2} \int_0^x dx' \left(\frac{x}{x'}\right)^2 \sin\left(\frac{x(x-x')K_\perp^2}{2kx'}\right) \hat{\varepsilon}_{\mathrm{mov}}\left(x', \frac{x\mathbf{K}_\perp}{x'}\right), \tag{19}$$

$$\hat{\phi}(x,\mathbf{K}_\perp) = \frac{k}{2} \int_0^x dx' \left(\frac{x}{x'}\right)^2 \cos\left(\frac{x(x-x')K_\perp^2}{2kx'}\right) \hat{\varepsilon}_{\mathrm{mov}}\left(x', \frac{x\mathbf{K}_\perp}{x'}\right). \tag{20}$$

Comparison between these formulas reveals that the only difference between $\hat{\chi}$ and $\hat{\phi}$ is that the integrand in Eq. (19)

contains the sine function while that in Eq. (20) contains cosine of the same argument.

Equations (19) and (20) for $\hat{\chi}$ and $\hat{\phi}$ are different from those presented in Ref. 11 for the case of plane wave propagation [see Eqs. (22) and (23) from Ref. 11]. The latter equations can be obtained from Eqs. (19) and (20) if the ratio $(x/x')^2$ in the integrands of these equations and the ratio x/x' in the arguments of $\hat{\varepsilon}_{\mathrm{mov}}$, sin and cos are set to 1.

III. CORRELATION FUNCTIONS AND VARIANCES OF LOG-AMPLITUDE AND PHASE FLUCTUATIONS

In this section, we derive formulas for the correlation functions and variances of log-amplitude and phase fluctuations of a spherical sound wave.

A. General formulas

Let the two receivers shown in Fig. 1 be located at the points $(x,\mathbf{r}_1)$ and $(x,\mathbf{r}_2)$, i.e., in a plane perpendicular to the x-axis. The transverse correlation functions of log-amplitude and phase fluctuations at these points are defined as

$$B_\chi(x;\mathbf{r}_1,\mathbf{r}_2)=\langle\chi(x,\mathbf{r}_1)\chi(x,\mathbf{r}_2)\rangle,$$

$$B_\phi(x;\mathbf{r}_1,\mathbf{r}_2)=\langle\phi(x,\mathbf{r}_1)\phi(x,\mathbf{r}_2)\rangle. \tag{21}$$

Here, the brackets $\langle\ \rangle$ denote ensemble or time average. In Eq. (21), we express $\chi(x,\mathbf{r})$ and $\phi(x,\mathbf{r})$ as the Fourier integrals (15) and (16):

$$B_\chi(x;\mathbf{r}_1,\mathbf{r}_2)=\int d\mathbf{K}_{\perp,1}\int d\mathbf{K}_{\perp,2}\exp(i\mathbf{K}_{\perp,1}\cdot\mathbf{r}_1+i\mathbf{K}_{\perp,2}\cdot\mathbf{r}_2)\times\langle\hat{\chi}(x,\mathbf{K}_{\perp,1})\hat{\chi}(x,\mathbf{K}_{\perp,2})\rangle, \tag{22}$$

$$B_\phi(x,\mathbf{r}_1,\mathbf{r}_2)=\int d\mathbf{K}_{\perp,1}\int d\mathbf{K}_{\perp,2}\exp(i\mathbf{K}_{\perp,1}\cdot\mathbf{r}_1+i\mathbf{K}_{\perp,2}\cdot\mathbf{r}_2)\times\langle\hat{\phi}(x,\mathbf{K}_{\perp,1})\hat{\phi}(x,\mathbf{K}_{\perp,2})\rangle. \tag{23}$$

Substituting the value of $\hat{\chi}(x,\mathbf{K}_\perp)$ from Eq. (19) into Eq. (22), we obtain a formula for the correlation function of the log-amplitude fluctuations

$$B_\chi(x;\mathbf{r}_1,\mathbf{r}_2)=\frac{k^2}{4}\int_0^x dx_1\int_0^x dx_2\int d\mathbf{q}_1\int d\mathbf{q}_2\times\exp(i\mathbf{K}_{\perp,1}\cdot\mathbf{r}_1+i\mathbf{K}_{\perp,2}\cdot\mathbf{r}_2)\times\frac{x^4}{x_1^2x_2^2}\sin\left(\frac{x(x-x_1)K_{\perp,1}^2}{2kx_1}\right)\sin\left(\frac{x(x-x_2)K_{\perp,2}^2}{2kx_2}\right)\times\langle\hat{\varepsilon}_{\mathrm{mov}}(x_1,x\mathbf{K}_{\perp,1}/x_1)\hat{\varepsilon}_{\mathrm{mov}}(x_2,x\mathbf{K}_{\perp,2}/x_2)\rangle. \tag{24}$$

Here, x_1 and x_2 are the integration variables along the x axis. In Eq. (24), let us change the integration variables $\mathbf{K}_{\perp,1}$ and $\mathbf{K}_{\perp,2}$ to new ones: $\mathbf{q}_1=(x/x_1)\mathbf{K}_{\perp,1}$ and $\mathbf{q}_2=(x/x_2)\mathbf{K}_{\perp,2}$. As a result, we have

$$B_\chi(x;\mathbf{r}_1,\mathbf{r}_2)=\frac{k^2}{4}\int_0^x dx_1\int_0^x dx_2\int d\mathbf{q}_1\int d\mathbf{q}_2\times\exp\left(i\mathbf{q}_1\cdot\mathbf{r}_1\frac{x_1}{x}+i\mathbf{q}_2\cdot\mathbf{r}_2\frac{x_2}{x}\right)\times\sin\left(\frac{x_1(x-x_1)q_1^2}{2kx}\right)\sin\left(\frac{x_2(x-x_2)q_2^2}{2kx}\right)\times\langle\hat{\varepsilon}_{\mathrm{mov}}(x_1,\mathbf{q}_1)\hat{\varepsilon}_{\mathrm{mov}}(x_2,\mathbf{q}_2)\rangle. \tag{25}$$

A formula for the correlation function of phase fluctuations can be derived similarly:

$$B_\phi(x;\mathbf{r}_1,\mathbf{r}_2)=\frac{k^2}{4}\int_0^x dx_1\int_0^x dx_2\int d\mathbf{q}_1\int d\mathbf{q}_2\times\exp\left(i\mathbf{q}_1\cdot\mathbf{r}_1\frac{x_1}{x}+i\mathbf{q}_2\cdot\mathbf{r}_2\frac{x_2}{x}\right)\times\cos\left(\frac{x_1(x-x_1)q_1^2}{2kx}\right)\cos\left(\frac{x_2(x-x_2)q_2^2}{2kx}\right)\times\langle\hat{\varepsilon}_{\mathrm{mov}}(x_1,\mathbf{q}_1)\hat{\varepsilon}_{\mathrm{mov}}(x_2,\mathbf{q}_2)\rangle. \tag{26}$$

The second moment $\langle\hat{\varepsilon}_{\mathrm{mov}}(x_1,\mathbf{q}_1)\hat{\varepsilon}_{\mathrm{mov}}(x_2,\mathbf{q}_2)\rangle$ appearing in Eqs. (25) and (26) was calculated in Ref. 11 [see Eq. (16) from that reference]:

$$\langle\hat{\varepsilon}_{\mathrm{mov}}(x_1,\mathbf{q}_1)\hat{\varepsilon}_{\mathrm{mov}}(x_2,\mathbf{q}_2)\rangle=\frac{\delta(x_1-x_2)}{2\pi}\int d\mathbf{r}_+e^{-i(\mathbf{q}_1+\mathbf{q}_2)\mathbf{r}_+}\Phi_{\mathrm{eff}}\left(x_1,\mathbf{r}_+;0,\frac{\mathbf{q}_1-\mathbf{q}_2}{2}\right). \tag{27}$$

Here, $\delta(x)$ is the delta function, and Φ_{eff} is the effective three-dimensional spectral density:

$$\Phi_{\mathrm{eff}}(x_+,\mathbf{r}_+;K_x,\mathbf{K}_\perp)=\frac{\Phi_T(x_+,\mathbf{r}_+;K_x,\mathbf{K}_\perp)}{T_0^2}+\frac{4\Phi_{xx}(x_+,\mathbf{r}_+;K_x,\mathbf{K}_\perp)}{c_0^2}, \tag{28}$$

where K_x is the x-component of the turbulence wave vector $\mathbf{K}=(K_x,\mathbf{K}_\perp)$, and Φ_T and Φ_{xx} are the three-dimensional spectral densities of the random temperature and velocity fields $\tilde{T}$ and v_x for the case of inhomogeneous, anisotropic turbulence. The only approximation made in deriving Eq. (27) is the Markov approximation.

For simplicity, it is assumed in Eq. (28) that the random fields $\tilde{T}$ and v_x are uncorrelated. If there is a cross-correlation between these random fields, a formula for Φ_{eff} is given by Eq. (12) from Ref. 11. In Eq. (28), the dependence of the three-dimensional spectral densities Φ_{eff}, Φ_T, and Φ_{xx} on x_+ and $\mathbf{r}_+$ is due to the inhomogeneity of the turbulence. For the case of homogeneous turbulence, Φ_{eff}, Φ_T, and Φ_{xx} do not depend on x_+ and $\mathbf{r}_+$ so that Eq. (28) coincides with Eq. (7.43) from Ref. 1.

The spectral densities $\Phi_T(x_+,\mathbf{r}_+;K_x,\mathbf{K}_\perp)$ and $\Phi_{xx}(x_+,\mathbf{r}_+;K_x,\mathbf{K}_\perp)$ can be related to the correlation func-

tions of temperature and velocity fluctuations, $B_T(\mathbf{R}_1,\mathbf{R}_2)=\langle\tilde{T}(\mathbf{R}_1)\tilde{T}(\mathbf{R}_2)\rangle$ and $B_{11}(\mathbf{R}_1,\mathbf{R}_2)=\langle v_x(\mathbf{R}_1)v_x(\mathbf{R}_2)\rangle$, by using the following approach. First, let us introduce new coordinates $\mathbf{R}_+=(\mathbf{R}_1+\mathbf{R}_2)/2$ and $\mathbf{R}_-=\mathbf{R}_1-\mathbf{R}_2$, and express the vectors $\mathbf{R}_+$ and $\mathbf{R}_-$ as $\mathbf{R}_+=(x_+,\mathbf{r}_+)$ and $\mathbf{R}_-=(x_-,\mathbf{r}_-)$. Then, the correlation functions of temperature and velocity fluctuations can be written as: $B_T(x_+,\mathbf{r}_+;x_-,\mathbf{r}_-)$ and $B_{11}(x_+,\mathbf{r}_+;x_-,\mathbf{r}_-)$. Finally, these correlation functions are expressed as the three-dimensional Fourier integrals with respect to x_- and $\mathbf{r}_-$. The spectral densities in these Fourier integrals are the functions $\Phi_T(x_+,\mathbf{r}_+;K_x,\mathbf{K}_\perp)$ and $\Phi_{xx}(x_+,\mathbf{r}_+;K_x,\mathbf{K}_\perp)$ appearing in Eq. (28).

B. Derivation of formulas for the correlation functions

In Eqs. (25) and (26), let us introduce new transverse coordinates:

$$\mathbf{r}_-=\mathbf{r}_1-\mathbf{r}_2,\quad \mathbf{r}_+=(\mathbf{r}_1+\mathbf{r}_2)/2, \tag{29}$$

and denote the correlation functions of log-amplitude and phase fluctuations as $B_\chi(x;\mathbf{r}_+,\mathbf{r}_-)$ and $B_\phi(x;\mathbf{r}_+,\mathbf{r}_-)$. The coordinates $\mathbf{r}_-$ and $\mathbf{r}_+$ characterize the distance between two receivers and their geometrical center in a plane $x=\text{const}$. Substituting Eq. (27) into Eq. (25) and calculating the integral over x_2, we have

$$\begin{aligned}B_\chi(x;\mathbf{r}_+,\mathbf{r}_-)&=\frac{k^2}{8\pi}\int_0^x dx_1\int d\mathbf{q}_1\int d\mathbf{q}_2\int d\mathbf{r}'\,\Phi_{\text{eff}}\left(x_1,\mathbf{r}';0,\frac{\mathbf{q}_1-\mathbf{q}_2}{2}\right)\\&\times\exp\left\{i\left[(\mathbf{q}_1+\mathbf{q}_2)\cdot\left(\mathbf{r}_+\frac{x_1}{x}-\mathbf{r}'\right)+(\mathbf{q}_1-\mathbf{q}_2)\cdot\mathbf{r}_-\frac{x_1}{2x}\right]\right\}\\&\times\sin\left(\frac{x_1(x-x_1)q_1^2}{2kx}\right)\sin\left(\frac{x_1(x-x_1)q_2^2}{2kx}\right).\end{aligned} \tag{30}$$

This equation is similar to Eq. (19-73) from Ref. 2. The difference between these equations is that the latter one was derived for the case of turbulence with temperature fluctuations only. Furthermore, in Eq. (17-97) from Ref. 2 the spectrum of temperature fluctuations $\Phi_T(x_+,\mathbf{r}_+;0,\mathbf{q})$ [which is similar to Φ_{eff} in Eq. (30)] was written as a product

$$\Phi_T(x_+,\mathbf{r}_+;0,\mathbf{q})=B_T^g(x_+,\mathbf{r}_+)\Phi_T^l(0,\mathbf{q}). \tag{31}$$

This formula describes, for example, the Kolmogorov spectrum of temperature fluctuations with the structure parameter depending on $(x_+,\mathbf{r}_+)$. However, Eq. (31) is not valid in a general case. For example, it is not valid if the outer scale length of turbulence depends on the height above the ground (as it does in the atmosphere). Therefore, Eq. (31) is not used in this paper.

In Eq. (30), the integration over $\mathbf{r}'$ is replaced by integration over a new vector $\mathbf{r}''=\mathbf{r}_+x_1/x-\mathbf{r}'$. As a result, we have

$$\begin{aligned}B_\chi(x;\mathbf{r}_+,\mathbf{r}_-)&=\frac{k^2}{8\pi}\int_0^x dx_1\int d\mathbf{q}_1\int d\mathbf{q}_2\int d\mathbf{r}''\Phi_{\text{eff}}\left(x_1,\mathbf{r}_+\frac{x_1}{x}-\mathbf{r}'';0,\frac{\mathbf{q}_1-\mathbf{q}_2}{2}\right)\exp\left\{i\left[(\mathbf{q}_1+\mathbf{q}_2)\cdot\mathbf{r}''+(\mathbf{q}_1-\mathbf{q}_2)\cdot\mathbf{r}_-\frac{x_1}{2x}\right]\right\}\\&\times\sin\left(\frac{x_1(x-x_1)q_1^2}{2kx}\right)\sin\left(\frac{x_1(x-x_1)q_2^2}{2kx}\right).\end{aligned} \tag{32}$$

In Eq. (32), the integration over the domains $q_1^2>k/x$ and $q_2^2>k/x$ contributes to B_χ only slightly since, in these domains, the sine functions oscillate rapidly. Therefore, the domains essential for integration over $\mathbf{q}_1$ and $\mathbf{q}_2$ are given by $q_1^2<k/x$ and $q_2^2<k/x$. Similarly, the domain essential for integration over $\mathbf{r}''$ is $r''<1/|\mathbf{q}_1+\mathbf{q}_2|<\sqrt{x/k}$. If the scale L_+ of $\Phi_{\text{eff}}(x_+,\mathbf{r}_+;0,\mathbf{q})$ with respect to $\mathbf{r}_+$ is greater than $\sqrt{x/k}$, the argument $\mathbf{r}_+x_1/x-\mathbf{r}''$ of Φ_{eff} in Eq. (32) can be replaced by $\mathbf{r}_+x_1/x$. In this case, the integral over $\mathbf{r}''$ in Eq. (32) is equal to $4\pi^2\delta(\mathbf{q}_1+\mathbf{q}_2)$. Then, calculating the integral over $\mathbf{q}_2$, we have

$$\begin{aligned}B_\chi(x,\mathbf{r}_+,\mathbf{r}_-)&=\frac{\pi k^2}{2}\int_0^x dx_1\int d\mathbf{q}_1\Phi_{\text{eff}}(x_1,\mathbf{r}_+x_1/x;0,\mathbf{q}_1)\\&\times\exp\left(i\mathbf{q}_1\cdot\mathbf{r}_-\frac{x_1}{x}\right)\sin^2\left(\frac{x_1(x-x_1)q_1^2}{2kx}\right).\end{aligned} \tag{33}$$

Introducing a new integration variable $\eta=x_1/x$, and denoting $K_F^2=k/x$ and $\mathbf{q}_1=\mathbf{K}_\perp$, we obtain the desired formula for the correlation function of log-amplitude fluctuations of a spherical sound wave propagating through inhomogeneous, anisotropic turbulence with temperature and velocity fluctuations:

$$\begin{aligned}B_\chi(x;\mathbf{r}_+,\mathbf{r}_-)&=\frac{\pi k^2x}{2}\int_0^1 d\eta\int d\mathbf{K}_\perp e^{i\eta\mathbf{K}_\perp\cdot\mathbf{r}_-}\sin^2\left(\frac{\eta(1-\eta)K_\perp^2}{2K_F^2}\right)\\&\times\Phi_{\text{eff}}(\eta x,\eta\mathbf{r}_+;0,\mathbf{K}_\perp).\end{aligned} \tag{34}$$

Here, $K_F=(\sqrt{x/k})^{-1}$, where a subscript "F" refers to the Fresnel zone.

Analogously, we obtain a formula for the correlation function of phase fluctuations of a spherical wave:

$$\begin{aligned}B_\phi(x;\mathbf{r}_+,\mathbf{r}_-)&=\frac{\pi k^2x}{2}\int_0^1 d\eta\int d\mathbf{K}_\perp e^{i\eta\mathbf{K}_\perp\cdot\mathbf{r}_-}\cos^2\left(\frac{\eta(1-\eta)K_\perp^2}{2K_F^2}\right)\\&\times\Phi_{\text{eff}}(\eta x,\eta\mathbf{r}_+;0,\mathbf{K}_\perp).\end{aligned} \tag{35}$$

In Eqs. (34) and (35), $r_-\ll x$ and $r_+\ll x$ since a parabolic approximation was used to derive these equations.

C. Analysis of formulas for the correlation functions

Formulas for B_χ and B_ϕ of a plane sound wave propagating through inhomogeneous, anisotropic turbulence are

given by Eqs. (29) and (30) from Ref. 11. These formulas coincide with Eqs. (34) and (35) in the present paper if the arguments of the exponential and trigonometric functions in the integrands are divided by η, and the argument $\eta\mathbf{r}_+$ of Φ_{eff} is replaced by $\mathbf{r}_+$. In two limiting cases, Eqs. (34) and (35) coincide with those known in the literature. First, Eqs. (34) and (35) coincide with Eqs. (7.68) and (7.69) from Ref. 1 for the case of homogeneous, isotropic turbulence when $\Phi_{\text{eff}}(x_+,\mathbf{r}_+;K_x,\mathbf{K}_\perp)$ does not depend on x_+ and $\mathbf{r}_+$. Second, Eqs. (34) and (35) with $\mathbf{r}_+=0$ can be obtained from Eq. (17-57) in Ref. 2 and Eq. (2.75) in Ref. 17 for the case of inhomogeneous, anisotropic turbulence with temperature fluctuations only.

The difference between B_χ and B_ϕ given by Eqs. (34) and (35) is that the former of these correlation functions contains the squared-sine function in the integrand while the latter contains the squared-cosine function of the same argument. Due to the squared-sine function, the scale of largest eddies which affect B_χ is of order $1/K_F$. On the other hand, due to the squared-cosine function, B_ϕ is affected by the largest eddies in the ABL. The scale $L_{\max}$ of these eddies can be of order of the height of the ABL. It should be noted that the Markov approximation, which was used to derive Eqs. (34) and (35), is valid if the distance x of sound propagation is greater than the scale of largest eddies which affect B_χ and B_ϕ. Therefore, the Markov approximation can always be used for calculations of B_χ since the requirement $x\gg 1/K_F$ is equivalent to $x\gg 1/k$. On the other hand, this approximation can be used for calculations of B_ϕ if $x\gg L_{\max}$. This inequality may not hold for some cases. In these cases, it is better to characterize phase fluctuations by the structure function $D_\phi(x;\mathbf{r}_+,\mathbf{r}_-)$ rather than by B_ϕ and $\langle\phi^2\rangle$. The Appendix presents derivation of a formula for the structure function of phase fluctuations of a spherical sound wave propagating through inhomogeneous anisotropic turbulence with temperature and velocity fluctuations [see Eq. (A6)].

Setting $\mathbf{r}_1=\mathbf{r}_2$ in Eqs. (34) and (35), we arrive at formulas for the variances of log-amplitude and phase fluctuations, $\langle\chi^2\rangle$ and $\langle\phi^2\rangle$, of a spherical sound wave propagating through inhomogeneous, anisotropic turbulence with temperature and velocity fluctuations:

$$\langle\chi^2\rangle = B_\chi(x;\mathbf{r}_+,0) = \frac{\pi k^2 x}{2}\int_0^1 d\eta\int d\mathbf{K}_\perp \sin^2\left(\frac{\eta(1-\eta)K_\perp^2}{2K_F^2}\right) \times\Phi_{\text{eff}}(\eta x,\eta\mathbf{r}_+;0,\mathbf{K}_\perp), \tag{36}$$

$$\langle\phi^2\rangle = B_\phi(x;\mathbf{r}_+,0) = \frac{\pi k^2 x}{2}\int_0^1 d\eta\int d\mathbf{K}_\perp \cos^2\left(\frac{\eta(1-\eta)K_\perp^2}{2K_F^2}\right) \times\Phi_{\text{eff}}(\eta x,\eta\mathbf{r}_+;0,\mathbf{K}_\perp). \tag{37}$$

Equations (34)–(37) for the correlation functions and variances of log-amplitude and phase fluctuations of a spherical sound wave propagating through inhomogeneous, anisotropic turbulence are the main analytical results obtained in this paper. Note that there is some freedom in choosing the x axis, which characterizes a predominant direction of sound propagation. When considering the correlation functions of log-amplitude and phase fluctuations, it is simplest to assume that the x axis passes through the geometrical center of two receivers. In this case, $\mathbf{r}_+=0$ in Eqs. (34)–(37) so that these equations simplify.

To study the effects of turbulence inhomogeneity and anisotropy on the correlation functions and variances of log-amplitude and phase fluctuations, one needs to specify the three-dimensional spectral densities of temperature and velocity fluctuations. Realistic spectra of velocity fluctuations in the atmosphere are considered in the next section.

IV. VELOCITY SPECTRA IN THE ABL

In this and the following sections, we consider the case when the source and the center mass of two receivers are located at the same height h above the ground (see Fig. 1). In this case, the x axis is parallel to the ground.

A. Inhomogeneous, anisotropic turbulence

A three-dimensional velocity spectral tensor $\Phi_{ij}(\mathbf{K})$ of turbulence produced by a constant shear in free space was obtained by Mann[18] with the use of rapid distortion theory. Here, $i=1,2,3$ and $j=1,2,3$ stand for components v_i and v_j of velocity fluctuations $\mathbf{v}$. Mann's result for $\Phi_{ij}(\mathbf{K})$ is anisotropic and has three free parameters: a variance σ_M^2, an outer length scale L_M, and an empirical constant Γ_M. The model can be applied heuristically to a surface-based shear layer by setting L_M proportional to the height h.[19] This assumption makes the model inhomogeneous, although in its original derivation the model is strictly homogeneous. Mann[18] also obtained a three-dimensional spectral tensor $\Phi_{ij}(h;\mathbf{K})$ for inhomogeneous, anisotropic turbulence by accounting for ground blocking with a more rigorous procedure. However, the resulting spectrum is complicated and difficult to use in calculations of sound propagation.

Adopting the heuristic approach to height dependence of the spectrum, Wilson[9] determined the free parameters in Mann's spectral model as follows:

$$\sigma_M^2=1.52\ u_*^2,\quad L_M=0.587\ h,\quad \Gamma_M=3.58, \tag{38}$$

where u_* is the friction velocity. Due to realistic dependence of L_M on h, Mann's spectrum with parameters given by Eq. (38) is a good model for a spectrum of inhomogeneous, anisotropic, shear-produced turbulence. This spectrum was used in Ref. 11 for numerical studies of the effects of turbulence inhomogeneity and anisotropy on plane wave sound propagation through a turbulent atmosphere. In Sec. V, this spectrum will be used in numerical studies of spherical wave propagation through inhomogeneous, anisotropic turbulence.

For the considered case of horizontal sound propagation through a turbulent atmosphere with Mann's spectrum, the effective three-dimensional spectral density $\Phi_{\text{eff}}(x_+,\mathbf{r}_+;0,\mathbf{K}_\perp)$ appearing in Eqs. (34)–(37) can be simplified. First, assuming that the x axis passes through the geometrical center of two receivers, we set $\mathbf{r}_+=0$. Second, Φ_{eff} does not depend on x_+ since turbulence is homogeneous in a horizontal plane. Finally, we take into account that Φ_{eff} parametrically depends on the height h above the ground. The result-

ing effective three-dimensional spectral density will be denoted as $\Phi_{\text{eff}}(h;0,\mathbf{K}_\perp)$. A formula for $\Phi_{\text{eff}}(h;0,\mathbf{K}_\perp)$ corresponding to Mann's spectrum is obtained in Ref. 11:

$$\Phi_{\text{eff}}(h;0,\mathbf{K}_\perp)=\frac{E(K_0)}{\pi c_0^2 K_0^4}\left\{K_y^2\left[1+\frac{K_0^4\cot^2\theta}{K_y^4}\times\arctan^2\left(-\frac{\beta K_y^2\sin\theta}{K_0^2-\beta K_{30}K_y\sin\theta}\right)\right]+K_{30}^2\right\}. \tag{39}$$

Here, θ is the azimuthal angle between the x axis and the direction of the mean horizontal wind, and the components of the turbulence wave vector $\mathbf{K}_0=(K_1,K_2,K_{30})$ are determined in terms of the components of the vector $\mathbf{K}_\perp=(K_y,K_z)$:

$$K_1=-K_y\sin\theta,\quad K_2=K_y\cos\theta,\qquad K_{30}=K_z-\beta K_y\sin\theta. \tag{40}$$

Furthermore, $K_0=|\mathbf{K}_0|$, and the initial energy spectrum of turbulence is given by

$$E(K_0)=\frac{55}{9B(\frac{1}{2},\frac{1}{3})}\frac{\sigma_M^2K_0^4L_M^5}{(1+K_0^2L_M^2)^{17/6}}, \tag{41}$$

where B is the beta function. In Eqs. (39) and (40), β is determined as

$$\beta=\frac{\sqrt{3}\Gamma_M}{KL_M}\left[B_{1/(1+K^2L_M^2)}\left(\frac{1}{3},\frac{5}{2}\right)\right]^{-1/2}, \tag{42}$$

where $B_\alpha(\frac{1}{3},\frac{5}{2})$ is the incomplete beta function, and K is the modulus of the vector $\mathbf{K}=(K_1,K_2,K_z)$.

B. Correlation functions and variances for Mann's spectrum

For the effective spectral density $\Phi_{\text{eff}}(h;0,\mathbf{K}_\perp)$ given by Eq. (39), the correlation functions and variances of log-amplitude and phase fluctuations can be written in a simplified form. Equation (39) is substituted into Eqs. (34) and (35). In the resulting formulas, we introduce a new integration vector $\overline{\mathbf{K}}_\perp=\mathbf{K}_\perp h$ and the wave parameter $D=x/(kh^2)$. As a result, we obtain formulas for the correlation functions of log-amplitude and phase fluctuations of a spherical sound wave for Mann's spectrum:

$$B_{\chi,\phi}(x;0,\mathbf{r}_-)=\frac{\pi k^2xhM^2}{4}\int_0^1 d\eta\int d\overline{\mathbf{K}}_\perp\frac{\Phi_{\text{eff}}(h;0,\overline{\mathbf{K}}_\perp)}{h^3M^2}\times e^{i\eta\overline{\mathbf{K}}_\perp\cdot\mathbf{r}_-/h}[1\mp\cos(\overline{\mathbf{K}}_\perp^2D\eta(1-\eta))]. \tag{43}$$

Here, the minus and plus signs correspond to B_χ and B_ϕ, respectively, and $M=u_*/c_0$ is a Mach number for the turbulent velocity fluctuations. Formulas for B_χ and B_ϕ of a plane sound wave can be obtained from Eq. (43) by dividing the arguments of the exponential and cosine functions by η.

In Eq. (43), the ratio $\Phi_{\text{eff}}/(h^3M^2)$ depends only on the wave parameter D. [This ratio also depends on σ_M^2/u_*^2 and L_M/h, which, however, are constants for Mann's spectrum; see Eq. (38).] Therefore, B_χ and B_ϕ given by Eq. (43) are functions of D, k^2xhM^2, and $\mathbf{r}_-/h$.

Setting $\mathbf{r}_-=0$ in Eq. (43) results in formulas for the variances of log-amplitude and phase fluctuations of a spherical sound wave for Mann's spectrum:

$$\langle\chi^2\rangle,\langle\phi^2\rangle=\frac{\pi k^2xhM^2}{4}\int_0^1 d\eta\int d\overline{\mathbf{K}}_\perp\frac{\Phi_{\text{eff}}(h;0,\overline{\mathbf{K}}_\perp)}{h^3M^2}\times[1\mp\cos(\overline{\mathbf{K}}_\perp^2D\eta(1-\eta))]. \tag{44}$$

It follows from this formula that the normalized variances, $\langle\chi^2\rangle/(k^2xhM^2)$ and $\langle\phi^2\rangle/(k^2xhM^2)$, depend only on the wave parameter D. In the particular case of Eq. (44), the integral over η can be calculated analytically, because Φ_{eff} does not depend on η. However, in other situations (such as a slanted propagation path) Φ_{eff} depends on η [see Eqs. (36) and (37)].

C. Inhomogeneous, isotropic turbulence

To better understand the effects of turbulence anisotropy on sound propagation, it is worthwhile to compare the correlation functions and variances of log-amplitude and phase fluctuations given by Eqs. (43) and (44) with those for the case of sound propagation through isotropic turbulence. Such isotropic turbulence will be modeled by von the Kármán spectrum of velocity fluctuations:[1,20]

$$\Phi_{ij}(h;\mathbf{K})=\frac{\Gamma(\frac{17}{6})}{\pi^{3/2}\Gamma(\frac{1}{3})}\frac{\sigma_v^2L_v^5(K^2\delta_{ij}-K_iK_j)}{(1+K^2L_v^2)^{17/6}}. \tag{45}$$

Here, Γ is the gamma-function, $\delta_{i,j}$ is the Kronecker symbol, σ_v^2 is the variance of one of the velocity components, and L_v is an outer scale of the velocity fluctuations. Using turbulence similarity theories, the following formulas for σ_v^2 and L_v were obtained for shear-produced turbulence:[9] $\sigma_v^2=3.0u_*^2$ and $L_v=1.8\,h$. Due to the dependence of L_v on h, the isotropic spectrum (45) is inhomogeneous.

Let us compare Mann's spectrum with the von Kármán spectrum. In the inertial subrange [$K_0\gg L_M$ in Eq. (41) and $K\gg L_v$ in Eq. (45)], where the turbulence is isotropic, Mann's spectrum Φ_{ij} with parameters given by Eq. (38) coincides with the von Kármán spectrum given by Eq. (45). On the other hand, in the energy subrange [$K_0\ll L_M$ in Eq. (41) and $K\ll L_v$ in Eq. (45)], Mann's spectrum is anisotropic and differs from the von Kármán spectrum which is isotropic. Therefore, if the statistical moments of a sound field are mainly affected by turbulence in the inertial subrange, they should be nearly the same for Mann's spectrum and for the von Kármán spectrum. However, if these statistical moments are mainly affected by turbulence in the energy subrange, they should be different for these two spectra. Also note that, in the inertial subrange, both the Mann and von Kármán spectra coincide with the Kolmogorov spectrum which is a realistic spectrum of homogeneous and isotropic turbulence. However, in the energy subrange, the former two spectra differ from the Kolmogorov spectrum and more realistically capture some features of turbulence in this subrange.

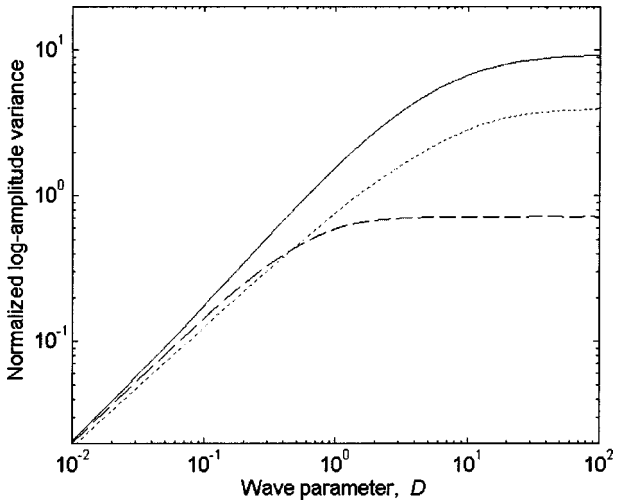


FIG. 2. Normalized log-amplitude variances, $\langle \chi^2 \rangle / (k^2 x h M^2)$, calculated for Mann's spectrum (solid line—along wind sound propagation; dashed line—crosswind) and an isotropic von Kármán spectrum (dotted line), as functions of the wave parameter $D = x/(kh^2)$.

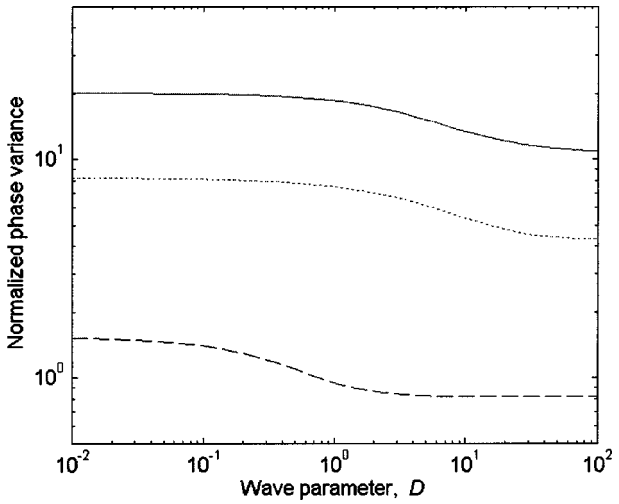


FIG. 3. Same as Fig. 2, but for the normalized phase variances, $\langle \phi^2 \rangle / (k^2 x h M^2)$.

The effective three-dimensional spectral density $\Phi_{\text{eff}}(h;0,\mathbf{K}_\perp)$ corresponding to the von Kármán spectrum (45) is given by

$$\Phi_{\text{eff}}(h;0,\mathbf{K}_\perp) = \frac{4}{c_0^2}\Phi_{11}(h;0,K_\perp) = \frac{4\Gamma(\frac{17}{6})}{\pi^{3/2}\Gamma(\frac{1}{3})c_0^2}\frac{\sigma_v^2 L_v^5 K_\perp^2}{(1+K_\perp^2 L_v^2)^{17/6}}. \quad (46)$$

Substitution of Eq. (46) into Eqs. (34)–(37) results in formulas for B_χ, B_ϕ, $\langle \chi^2 \rangle$, and $\langle \phi^2 \rangle$ for the von Kármán spectrum.

V. NUMERICAL RESULTS

In this section, we present results of numerical calculation of the correlation functions and variances of log-amplitude and phase fluctuations given by Eqs. (43) and (44). The trapezoidal method was used for all of the numerical integrations in these equations. In the numerical integrations, the η axis was partitioned into 30 linearly spaced points between 0 and 1. Furthermore, the two-dimensional integration over $\overline{\mathbf{K}}_\perp = (\overline{K}_y, \overline{K}_z)$ was performed over the region $-10^3 \leq \overline{K}_y \leq 10^3$, $-10^3 \leq \overline{K}_z \leq 10^3$. The $\overline{K}_y$ and $\overline{K}_z$ axes both had 400 points, spaced linearly at wavenumber magnitudes below 0.1 and logarithmically above.

A. Variances of log-amplitude and phase fluctuations

Normalized variances of log-amplitude and phase fluctuations, $\langle \chi^2 \rangle / (k^2 x h M^2)$ and $\langle \phi^2 \rangle / (k^2 x h M^2)$, are shown in Figs. 2 and 3 versus the wave parameter D. The solid and dashed lines correspond to the normalized variances for sound propagation through anisotropic turbulence with Mann's spectrum along and across the mean wind, whereas the dotted line corresponds to the case of isotropic turbulence with the von Kármán spectrum. It follows from Figs. 2 and 3 that, for along-wind propagation, the normalized variance of log-amplitude fluctuations for $D > 1$ and the normalized variance of phase fluctuations for all values of D are significantly greater than those for crosswind propagation, with the variances for isotropic turbulence falling in between. This result is due to the fact that, in this case, the variances are affected by turbulent eddies in the energy subrange, which are stretched in the direction of the mean wind. Therefore, the outer scale of turbulence and, hence, the log-amplitude and phase variances are greater in the along-wind than in the crosswind direction.

On the other hand, for $D < 1$, the variance of log-amplitude fluctuations is affected by turbulence in the inertial subrange (e.g., Ref. 5). Since Mann's spectrum coincides with the von Kármán spectrum in the inertial subrange, all three curves in Fig. 2 are close to each other for small values of D.

In Figs. 2 and 3, the normalization factor $k^2 x h M^2$ and the wave parameter $D = x/(kh^2)$ depend on the height h of the sound propagation path. Therefore, the variances of log-amplitude and phase fluctuations, $\langle \chi^2 \rangle$ and $\langle \phi^2 \rangle$, also depend on h. This demonstrates the effect of turbulence inhomogeneity on $\langle \chi^2 \rangle$ and $\langle \phi^2 \rangle$.

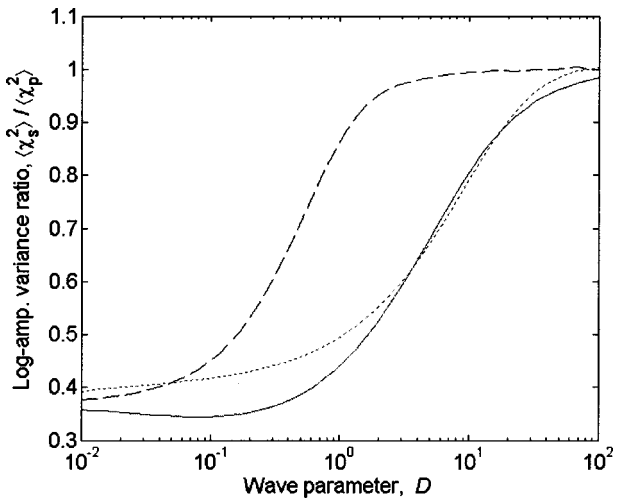


FIG. 4. Same as Fig. 2, except that the ratio of the log-amplitude variance for spherical waves, $\langle \chi_s^2 \rangle$, to that for plane waves, $\langle \chi_p^2 \rangle$, is shown.

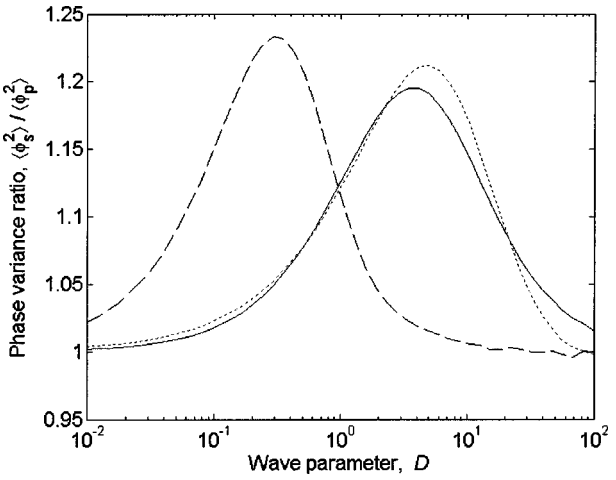


FIG. 5. Same as Fig. 2, except that the ratio of the phase variance for spherical waves, $\langle \phi_s^2 \rangle$, to that for plane waves, $\langle \phi_p^2 \rangle$, is shown.

The normalized variances of log-amplitude and phase fluctuations shown in Figs. 2 and 3 are very similar to those for plane wave propagation, considered in Ref. 11. Figures 4 and 5 show ratios of the variances for spherical waves to those for plane waves. The main difference is that, for $D \ll 1$, the variance of log-amplitude fluctuations for a spherical sound wave is less than that for a plane wave by a factor 2.5. This factor has been well established by theoretical analyses of sound propagation through inertial-subrange turbulence (e.g., Ref. 1, p. 206). It is also notable that the impact of anisotropy on the variances for spherical and plane waves is nearly the same when $D \ll 1$ or $D \gg 1$; the primary differences occur for moderate values of D.

An experiment described in Ref. 21 provides some qualitative evidence in support of the variance calculations. In the experiment, $\langle \chi^2 \rangle$ was measured as a function of frequency and propagation distance in two different conditions, one roughly crosswind and the other downwind. The mean wind speeds were comparable at 2.5 and 3.4 m/s, respectively. Saturation of $\langle \chi^2 \rangle$ with increasing distance was observed for the crosswind propagation, but not for the downwind propagation. This is consistent with Fig. 2, since D is proportional to x. Note that, in the experiment, the propagation path was slanted whereas in Secs. IV and V it is assumed to be horizontal. Therefore, Fig. 2 cannot be used for quantitative comparison between experimental data and theoretical predictions.

In order to experimentally verify theoretical results presented in Figs. 2 and 3, microphones installed in the direction of the mean wind and in the perpendicular direction should be used to measure the variances. As far as we know such experiments have not been previously completed and analyzed. It would be very worthwhile to carry out such experiments.

B. Correlation functions of log-amplitude and phase fluctuations

Normalized correlation functions of log-amplitude and phase fluctuations, $B_\chi(x;0,\mathbf{r}_-)/B_\chi(x;0,0)$ and $B_\phi(x;0,\mathbf{r}_-)/B_\phi(x;0,0)$, are plotted in Figs. 6–9 versus the normalized separation r_-/h between two receivers. The solid and dashed lines correspond to the normalized correlation functions for along-wind and crosswind sound propagation through anisotropic turbulence with Mann's spectrum. The dotted line corresponds to isotropic turbulence with the von Kármán spectrum. For along-wind and crosswind sound propagation, squares on the corresponding lines indicate vertical separation between receivers; circles indicate horizontal receiver separation. In Figs. 6 and 7, the wave parameter $D = 0.1$, whereas in Figs. 8 and 9, $D = 10$.

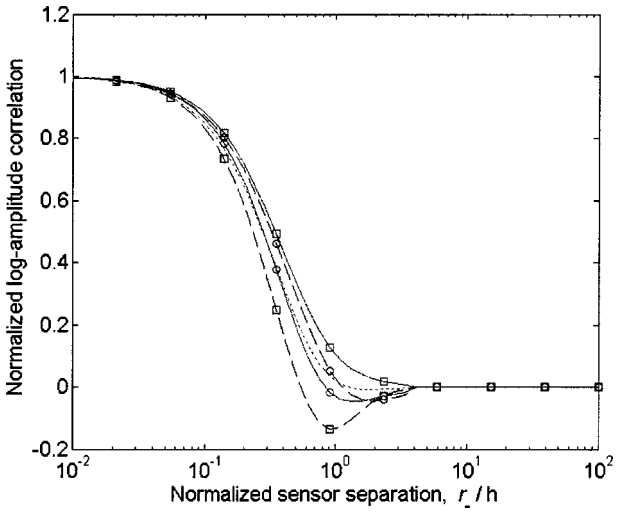


FIG. 6. Normalized correlation functions of log-amplitude fluctuations, $B_\chi(x;0,\mathbf{r}_-)/B_\chi(x;0,0)$, calculated for Mann's spectrum (solid line—along wind sound propagation; dashed line—crosswind) and an isotropic von Kármán spectrum (dotted line), as functions of the normalized receiver separation r_-/h. Squares indicate vertical receiver separations; circles indicate horizontal separations. The diffraction parameter is $D = 0.1$.

The normalized correlation functions of log-amplitude and phase fluctuations shown in Figs. 6–9 generally decrease as r_-/h increases. They are consistently near zero for $r_-/h \gg 1$. It follows from the figures that, for along-wind and crosswind sound propagation, the values of the correlation functions depend significantly on whether two receivers are separated in a vertical or horizontal plane. Furthermore, it

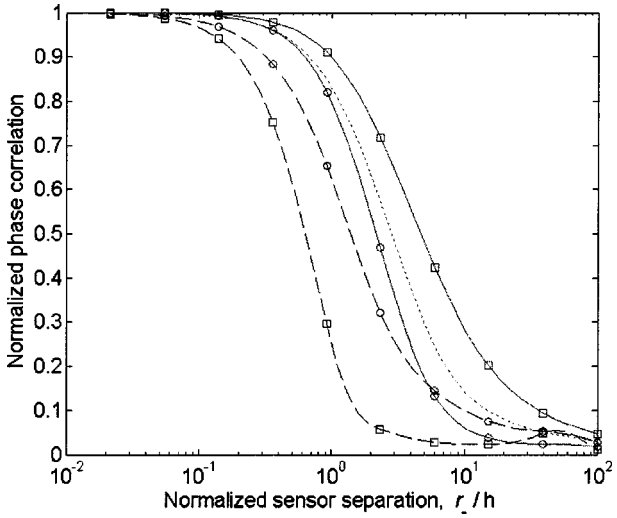


FIG. 7. Same as Fig. 6, but for the normalized correlation functions of phase fluctuations, $B_\phi(x;0,\mathbf{r}_-)/B_\phi(x;0,0)$.

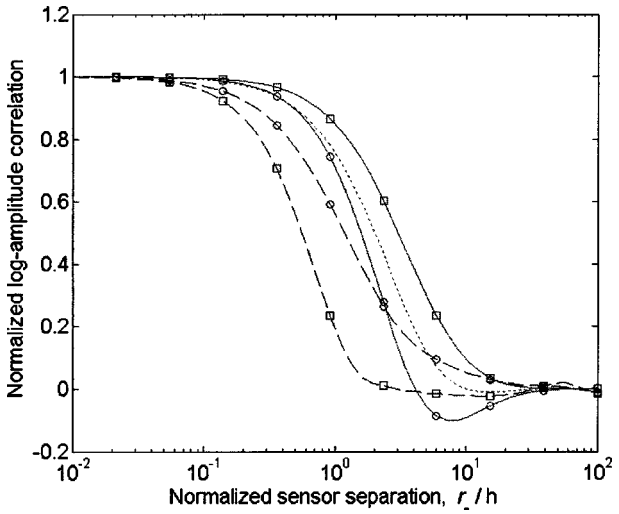


FIG. 8. Same as Fig. 6, but for $D=10$.

follows from Figs. 6–9 that, in most cases, the log-amplitude and phase correlations in the along-wind direction are greater than those in the crosswind direction. These results clearly show the effect of turbulence anisotropy on the normalized correlation functions. It should be noted here that, if the normalized vertical sensor separation is $r_-/h=2$, the lower sensor is located on the ground. Therefore, in Figs. 6–9, the lines with squares, strictly speaking, should be plotted in the range $0\leqslant r_-/h\leqslant 2$. Nevertheless, these lines are plotted in a wider range for a better visual comparison with other lines.

In Figs. 6–9 the normalized separation r_-/h depends on the height h of the sound propagation path above the ground. Therefore, the correlation functions of log-amplitude and phase fluctuations also depend on h. This dependence illustrates the effect of turbulence inhomogeneity on sound propagation.

It can be seen from Figs. 6 and 8 that the normalized correlation functions of log-amplitude fluctuations increase as the wave parameter D increases from 0.1 to 10. On the other hand, the normalized correlation functions of phase fluctuations depend on D only slightly (Figs. 7 and 9).

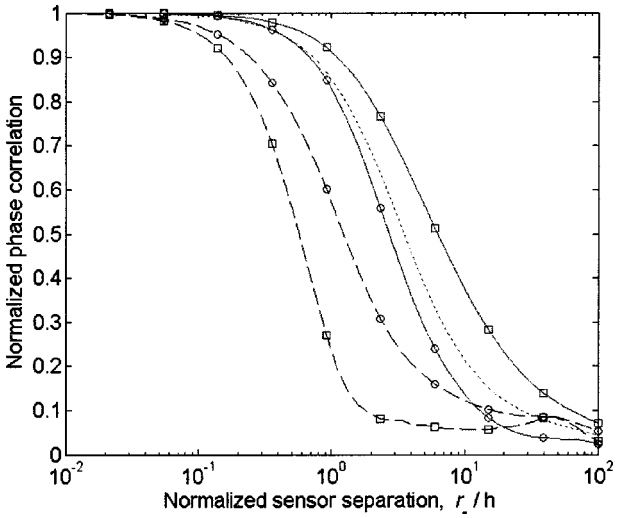


FIG. 9. Same as Fig. 6, but for the normalized correlation functions of phase fluctuations, $B_\phi(x;0,\mathbf{r}_-)/B_\phi(x;0,0)$, and $D=10$.

The normalized correlation functions shown in Figs. 6–9 look similar to those obtained in Ref. 11 for the case of plane wave propagation. The main difference between the correlation functions for spherical and plane wave propagation is that, in the former case, the correlation persists out to longer receiver separations. This can be understood from the fact that the spherical wave diverges as it propagates. The direct paths along which sound energy propagates from a source to two receivers are close to each other near the source and therefore initially pass through the same turbulent eddies, thereby minimizing loss of correlation.

VI. CONCLUSIONS

In this paper, we have studied log-amplitude and phase fluctuations of a spherical sound wave propagating through inhomogeneous, anisotropic turbulence with temperature and velocity fluctuations. Starting from the parabolic equation and using the Rytov method and the Markov approximation, we derived formulas for the correlation functions and variances of log-amplitude and phase fluctuations of this spherical wave. The Rytov method allows one to correctly describe weak fluctuations in log-amplitude of a sound wave.

These formulas generalize those known in the literature for two limiting cases: (a) homogeneous, isotropic turbulence, and (b) inhomogeneous, anisotropic turbulence with temperature fluctuations only. The formulas for the correlation functions and variances of log-amplitude and phase fluctuations derived in the present paper differ from those obtained in Ref. 11 for the case of plane wave propagation through inhomogeneous, anisotropic turbulence.

Using the derived formulas and Mann's spectral tensor of shear-driven turbulence, the correlation functions and variances of log-amplitude and phase fluctuations of a spherical sound wave have been studied numerically. The results obtained clearly show that turbulence inhomogeneity and anisotropy can significantly affect these statistical moments of a spherical sound wave.

ACKNOWLEDGMENTS

This material is partly based upon work supported by the U.S. Army Research Office under Contract No. DAAG19-01-1-0640 (administered by W. Bach).

APPENDIX: STRUCTURE FUNCTION OF PHASE FLUCTUATIONS OF A SPHERICAL SOUND WAVE

In this Appendix, we derive a formula for the structure function of phase fluctuations of a spherical sound wave propagating through inhomogeneous, anisotropic turbulence with temperature and velocity fluctuations.

The transverse structure function of phase fluctuations is defined as

$$D_\phi(x;\mathbf{r}_1,\mathbf{r}_2)=\langle[\phi(x,\mathbf{r}_1)-\phi(x,\mathbf{r}_2)]^2\rangle. \tag{A1}$$

If $\langle\phi^2\rangle$ and B_ϕ exist, the structure function can be written as

$$D_\phi(x;\mathbf{r}_1,\mathbf{r}_2)=\langle\phi^2(x,\mathbf{r}_1)\rangle+\langle\phi^2(x,\mathbf{r}_2)\rangle - 2B_\phi(x;\mathbf{r}_1,\mathbf{r}_2). \tag{A2}$$

In this formula, we change to the transverse coordinates $\mathbf{r}_-$ and $\mathbf{r}_+$ given by Eq. (29):

$$D_\phi(x;\mathbf{r}_+,\mathbf{r}_-)=\langle\phi^2(x,\mathbf{r}_+ +\mathbf{r}_-/2)\rangle+\langle\phi^2(x,\mathbf{r}_+ -\mathbf{r}_-/2)\rangle - 2B_\phi(x;\mathbf{r}_+,\mathbf{r}_-). \tag{A3}$$

In Eq. (A3), B_ϕ and $\langle\phi^2\rangle$ are replaced by their values from Eqs. (35) and (37). As a result, we have

$$\begin{aligned} D_\phi(x;\mathbf{r}_+,\mathbf{r}_-) &= \frac{\pi k^2 x}{2}\int_0^1 d\eta\left\{\int d\mathbf{q}_1\cos^2\left(\frac{\eta(1-\eta)q_1^2}{2K_F^2}\right)\right. \\ &\times\Phi_{\text{eff}}(\eta x,(\mathbf{r}_+ +\mathbf{r}_-/2)\eta;0,\mathbf{q}_1)+\int d\mathbf{q}_2 \\ &\times\cos^2\left(\frac{\eta(1-\eta)q_2^2}{2K_F^2}\right)\Phi_{\text{eff}}(\eta x,(\mathbf{r}_+ -\mathbf{r}_-/2)\eta;0,\mathbf{q}_2) \\ &-2\int d\mathbf{q}_3\exp(i\eta\mathbf{q}_3\cdot\mathbf{r}_-)\cos^2\left(\frac{\eta(1-\eta)q_3^2}{2K_F^2}\right) \\ &\left.\times\Phi_{\text{eff}}(\eta x,\eta\mathbf{r}_+;0,\mathbf{q}_3)\right\}. \end{aligned} \tag{A4}$$

In Sec. III B, it was assumed that $\Phi_{\text{eff}}(x,\mathbf{r}_+;0,\mathbf{q})$ varies slowly with respect to the argument $\mathbf{r}_+$. Therefore, the following approximation holds:

$$\Phi_{\text{eff}}(x,\mathbf{r}_+ +\mathbf{r}_-;0,\mathbf{q})=\Phi_{\text{eff}}(x,\mathbf{r}_+;0,\mathbf{q}) +\mathbf{r}_-\cdot\frac{\partial\Phi_{\text{eff}}(x,\mathbf{r}_+;0,\mathbf{q})}{\partial\mathbf{r}_+}. \tag{A5}$$

Using this approximation in the integrals over $\mathbf{q}_1$ and $\mathbf{q}_2$ in Eq. (A4), we obtain a desired formula for the structure function of phase fluctuations of a spherical sound wave propagating through inhomogeneous, anisotropic turbulence with temperature and velocity fluctuations:

$$D_\phi(x;\mathbf{r}_+,\mathbf{r}_-)=\pi k^2 x\int_0^1 d\eta\int d\mathbf{K}_\perp[1-e^{i\eta\mathbf{K}_\perp\cdot\mathbf{r}_-}] \times\cos^2\left(\frac{\eta(1-\eta)K_\perp^2}{2K_F^2}\right)\Phi_{\text{eff}}(\eta x,\eta\mathbf{r}_+;0,\mathbf{K}_\perp). \tag{A6}$$

This equation can also be derived by a more involved approach which, however, does not require that $\langle\phi^2\rangle$ and B_ϕ exist. In this approach, ϕ in Eq. (A1) is expressed in terms of $\hat{\varepsilon}_{\text{mov}}(x,\mathbf{K}_\perp)$ using Eqs. (16) and (20). The resulting formula contains second moments $\langle\hat{\varepsilon}_{\text{mov}}(x_1,\mathbf{q}_1)\hat{\varepsilon}_{\text{mov}}(x_2,\mathbf{q}_2)\rangle$ which are given by Eq. (27). Finally, assuming that the scale L_+ of variation of $\Phi_{\text{eff}}(x_+,\mathbf{r}_+;0,\mathbf{K}_\perp)$ with respect to $\mathbf{r}_+$ is greater than $\sqrt{x/k}$, we arrive at Eq. (A6).

Due to the factor $1-e^{i\eta\mathbf{K}_\perp\cdot\mathbf{r}_-}$ in the integrand in Eq. (A6), the scale of the largest eddies which contribute to D_ϕ is of order r_-. Since in the parabolic equation approximation x is always greater than r_-, Eq. (A6) can be used for calculations of D_ϕ whether or not x is greater than the scale $L_{\max}$ of largest eddies in the ABL. On the other hand, the range of applicability of Eqs. (35) and (37) for B_ϕ and $\langle\phi^2\rangle$ is limited to $x\gg L_{\max}$.

[1] V. E. Ostashev, *Acoustics in Moving Inhomogeneous Media* (E&FN SPON, London, 1997).
[2] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978).
[3] S. M. Flatté, R. Dashen, W. H. Munk, K. M. Watson, and F. Zachariasen, *Sound Transmission through a Fluctuating Ocean* (Cambridge U. P., New York, 1979).
[4] L. A. Chernov, *Wave Propagation in a Random Medium* (Dover, New York, 1960).
[5] S. M. Rytov, Y. A. Kravtsov, and V. I. Tatarskii, *Principles of Statistical Radio Physics. Part 4, Wave Propagation through Random Media* (Springer, Berlin, 1989).
[6] V. I. Tatarskii, *The Effects of the Turbulent Atmosphere on Wave Propagation* (Keter, Jerusalem, 1971).
[7] V. A. Krasil'nikov, "Linear and nonlinear sound propagation in turbulent and inhomogeneous media," Acoust. Phys. **44**, 481–490 (1998).
[8] D. K. Wilson and D. W. Thomson, "Acoustic propagation through anisotropic, surface-layer turbulence," J. Acoust. Soc. Am. **96**, 1080–1095 (1994).
[9] D. K. Wilson, "Calculated coherence and extinction of sound waves propagating through anisotropic, shear-induced turbulent velocity fluctuations," J. Acoust. Soc. Am. **105**, 658–671 (1999).
[10] M. Kelly, R. Raspet, and H. E. Bass, "Scattering of sonic booms by anisotropic turbulence in the atmosphere," J. Acoust. Soc. Am. **107**, 3059–3064 (2000).
[11] V. E. Ostashev and D. K. Wilson, "Log-amplitude and phase fluctuations of a plane wave propagating through anisotropic, inhomogeneous turbulence," Acust. Acta Acust. **87**, 685–694 (2001).
[12] Some of the results in this paper were presented previously in V. E. Ostashev, D. K. Wilson, and G. H. Goedecke, "Sound propagation and scattering through inhomogeneous, anisotropic atmospheric turbulence," in *Proceedings of Battlespace Atmospheric and Cloud Impacts on Military Operations Conference 2001*, Ft. Collins, CO, 2001.
[13] V. E. Ostashev, V. Mellert, R. Wandelt, and F. Gerdes, "Propagation of sound in a turbulent medium I. Plane waves," J. Acoust. Soc. Am. **102**, 2561–2570 (1997).
[14] V. E. Ostashev, F. Gerdes, V. Mellert, and R. Wandelt, "Propagation of sound in a turbulent medium II. Spherical waves," J. Acoust. Soc. Am. **102**, 2571–2578 (1997).
[15] S. L. Collier and D. K. Wilson, "Cramer-Rao lower bounds on the angle-of-arrival estimates for a wave propagating in a random medium: Geometric acoustics regime," J. Acoust. Soc. Am. **113**, 2312 (2003).
[16] S. L. Collier and D. K. Wilson, "Acoustic beamforming in atmospheric turbulence," 10th Long Range Sound Propagation Symposium, Grenoble, France (2002) (in print).
[17] A. S. Gurvich, A. I. Kon, V. L. Mironov, and S. S. Khmelevtsov, *Laser Radiation in the Turbulent Atmosphere* (in Russian) (Nauka, Moscow, 1976).
[18] J. Mann, "The spatial structure of neutral atmospheric surface layer turbulence," J. Fluid Mech. **273**, 141–168 (1994).
[19] J. Mann, "Wind field simulation," Prob. Eng. Mech. **13**, 269–282 (1998).
[20] D. K. Wilson and V. E. Ostashev, "A reexamination of acoustic scattering in the atmosphere using improved models for the turbulence spectrum," in *Proceedings of Battlespace Atmospheric and Cloud Impacts on Military Operations Conference 2000*, Ft. Collins, CO, 2000.
[21] D. K. Wilson, D. I. Havelock, M. Heyd, M. J. Smith, J. M. Noble, and H. J. Auvermann, "Experimental determination of the effective structure-function parameter for atmospheric turbulence," J. Acoust. Soc. Am. **105**, 912–914 (1999).

The sonar equation and the definitions of propagation loss

Michael A. Ainslie[a)]
TNO Physics and Electronics Laboratory, Underwater Acoustics Group, Oude Waalsdorperweg 63, 2509 JG The Hague, The Netherlands

(Received 24 July 2003; revised 10 October 2003; accepted 20 October 2003)

A rigorous application of the traditional definition of sonar equation terms leads to the appearance of an unexpected factor, not routinely included, equal to the ratio of the characteristic impedance at the receiver to that at the source. An omission of this factor can lead to non-negligible errors for realistic conditions. It is further argued that a gradual change in the *de facto* definition of propagation loss occurred between 1965 and 1980. Two alternatives to the traditional sonar equation are suggested, each using one of the two propagation loss definitions and both eliminating the unwanted impedance ratio. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1631939]

PACS numbers: 43.30.Wi [ADP] Pages: 131–134

I. INTRODUCTION

The sonar equation is a useful tool for quantifying the signal to noise ratio (snr) for the purpose of sonar performance prediction, usually either for analysis or design. The equation can take on different forms depending on the nature of the sonar (e.g., whether active or passive) and details of the processing such as narrow band or broadband. These forms are described in standard texts such as Urick's book.[1] In the case of a passive sonar, snr is usually defined as the ratio of the ''intensity'' of the signal arriving at the sonar receiver to that of the background noise. The quotation marks are necessary because this is not the true intensity of the acoustic signal itself, but that of an equivalent plane wave.

The purpose of this note is to point out some problems with the sonar equation expressed in this way and to show that these can be solved by adopting some practical alternatives. The definitions described as ''traditional'' are from Ref. 1, supplemented where necessary by clarifications from Ref. 2. Disregarding minor changes in terminology they are also consistent with Ref. 3 and have been restated recently in Ref. 4. Nevertheless, the definitions are not adopted unanimously, and Ref. 5 presents the sonar equation in terms of squared pressure rather than intensity. Several articles have appeared recently with the objective either of clarifying existing definitions[6,7] or of proposing new ones.[8] The appearance of Morfey's Dictionary of Acoustics[9] in 2001 should also be mentioned, as it includes rigorous definitions of some of the relevant terms.

The problems with the traditional definitions are explained at the end of Sec. II. Two alternative sonar equations are described in Sec. III, using slightly different definitions of the sonar equation terms. One of these alternatives, based on definitions proposed by Hall,[8] is consistent with the early propagation loss definition; the other uses the modern definition and is identical to the sonar equation used by Ross.[5] The distinction is rarely recognized, and one of the present objectives is to highlight the need for adopting a single, unambiguous, definition for propagation loss.

For simplicity the treatment here is for passive sonar, point sources, and time-stationary continuous signals. Stationarity implies that the energy per unit time (and its frequency spectrum) entering the receiving equipment is constant. An extension to transients, active sonar, directional sources, and hydrophone arrays is possible, but not necessary to meet the present objective. Issues regarding the use of the decibel are deliberately avoided by casting the equations in a linear rather than logarithmic form. Conclusions are drawn in Sec. IV.

II. THE TRADITIONAL SONAR EQUATION

The signal to noise ratio (snr) of an acoustic detection system can be written as a ratio of the signal intensity to the noise intensity: $\mathrm{snr} \equiv I_S^{\mathrm{eq}}/I_N^{\mathrm{eq}}$, where the subscripts ''*S*'' and ''*N*'' indicate *signal* and *noise*, respectively. The ''eq'' superscript indicates that this is not the true acoustic intensity (a vector), but an equivalent plane wave intensity, defined as the intensity magnitude of a plane progressive wave with the same root mean square (rms) pressure, and propagating in the same medium, as the true field.[2] This equivalent intensity is abbreviated henceforth as EPWI.

By convention the signal EPWI is characterized by the source level sl and propagation loss pl. The lower case notation is used to indicate that these are linear versions of the equivalent upper case quantities (SL and PL), usually expressed in dB. Here sl is the intensity I_0 of sound at a specified reference distance r_{ref} (usually 1 m) from the source, relative to a reference intensity I_{ref} (that of a plane wave of a specified rms pressure p_{ref}, usually 1 μPa):

$$\mathrm{sl} \equiv \frac{I_0}{I_{\mathrm{ref}}}, \tag{1}$$

and pl is the ratio of the source intensity I_0 to the signal EPWI:

$$\mathrm{pl} \equiv \frac{I_0}{I_S^{\mathrm{eq}}}. \tag{2}$$

a)Electronic mail: ainslie@fel.tno.nl

Thus, the ratio of these two quantities (sl/pl) gives the signal level $I_S^{\mathrm{eq}}/I_{\mathrm{ref}}$. Similarly, the noise level nl is the noise EPWI divided by the reference intensity,

$$\mathrm{nl} \equiv \frac{I_N^{\mathrm{eq}}}{I_{\mathrm{ref}}}, \tag{3}$$

and therefore

$$\mathrm{snr} = \frac{\mathrm{sl}}{\mathrm{pl} \cdot \mathrm{nl}}. \tag{4}$$

This last expression is the sonar equation in linear form, preferred here to its dB equivalent because it results in simpler algebraic expressions. The definitions, Eqs. (1)–(3), have been used by sonar designers and analysts for decades. Their expression in the form of simple dimensionless ratios makes them convenient for approximate engineering calculations, but it is argued below that some modifications are desirable for modern scientific work as follows:

A. Reference intensity and the sonar equation

The above definition of reference intensity amounts to $I_{\mathrm{ref}}(\mathbf{x}) \equiv p_{\mathrm{ref}}^2/Z(\mathbf{x})$, where $Z(\mathbf{x})$ is the characteristic impedance of the medium at the measurement location $\mathbf{x}$. It follows that I_{ref} is a function of position, so that Eq. (4) is only approximately true, and should be written instead as

$$\mathrm{snr} = \frac{\mathrm{sl}}{\mathrm{pl} \cdot \mathrm{nl}} \frac{I_{\mathrm{ref}}(\mathbf{x}_0)}{I_{\mathrm{ref}}(\mathbf{x})}, \tag{5}$$

where $\mathbf{x}_0$ denotes the source position. This may seem a minor point because variations in the impedance of seawater are usually small, and for many practical applications the correction is indeed negligible. Nevertheless, both density and sound speed vary with salinity, temperature, and pressure, and under realistic extremes of salinity (0–40 ppt) and pressure (1–500 atm) the impedance can differ by up to about 7% from a nominal value of 1.5 Mray1 ($1\,\mathrm{rayl} = 1\,\mathrm{kg\,m^{-2}\,s^{-1}}$). A 7% offset in linear units corresponds to a systematic error of 0.29 dB. Much larger changes are also possible, for example, in the presence of bubbles,[10] leading to correspondingly large errors from the use of Eq. (4)

B. Definition of propagation loss

The definition of Eq. (2) is not universally adopted for propagation loss pl. It was used for early ray tracing calculations based on computing ratios of areas between contiguous ray paths,[2,11,12] but an alternative convention prevalent in modern literature is to define pl as a ratio of mean square pressure (MSP) values. In other words,

$$\mathrm{pl}_{\mathrm{MSP}} \equiv p_0^2/\overline{p_S^2}, \tag{6}$$

where p_0 is the rms pressure that would be measured at a distance r_{ref} from an equivalent point source. This quantity is related to I_0 according to the equation $p_0 = \sqrt{I_0 Z(\mathbf{x}_0)}$, which means that the two pl definitions differ by a factor $Z(\mathbf{x}_0)/Z(\mathbf{x})$. The pl definition of Eq. (2) is henceforth referred to as $\mathrm{pl}_{\mathrm{EPWI}}$ to distinguish it from $\mathrm{pl}_{\mathrm{MSP}}$.

Early studies of propagation across an impedance boundary usually involved the underwater detection of a point source in air, some calculating EPWI,[2,11–13] some velocity potential[14,15] and others pressure.[16–18] The following discussion considers only those publications that use the term ''propagation loss'' (or ''transmission loss;'' no distinction is made between these two terms, which in this context are synonymous[9]).

Contours of $\mathrm{pl}_{\mathrm{EPWI}}$ from air to water and from water to air appear in a landmark publication by Horton.[2] The same definition is used later by Urick[11] and by Young[12] for their studies of propagation from air to water. Bordelon[18] makes a distinction between $\mathrm{pl}_{\mathrm{MSP}}$ and $\mathrm{pl}_{\mathrm{EPWI}}$ and seems to prefer the former. He converts Horton's EPWI values to MSP and compares them with his own MSP predictions. Deavenport and DiNapoli[19] use $\mathrm{pl}_{\mathrm{EPWI}}$ for their wave theory calculations of air-to-water propagation.

A number of articles considering propagation from a point source in water, into the sediment, started to appear in the early 1980s. The earliest of these[19] applies the EPWI definition, but subsequent articles (to this author's knowledge up until the date of publication) applied to this problem unanimously use MSP. Early examples are Refs. 20–22.

There are two more articles, both by Pedersen and Gordon,[23,24] that are relevant to this discussion, even though they do not explicitly consider propagation across a physical boundary. Reference 23 compares a ray theory result using $\mathrm{pl}_{\mathrm{EPWI}}$ with a wave theory computation using $\mathrm{pl}_{\mathrm{MSP}}$. Reference 24 then discusses the difference between the two definitions, seeming to consider EPWI the more fundamental.

Also relevant are textbook definitions. Although the EPWI definition is quoted in some standard texts,[1,4] the MSP definition is more common.[5,9,25,26] Even Ref. 4, despite insisting (p. 11) that ''The decibel ... denotes a ratio of intensities (not pressures)'' and (p. 12) ''transmission loss [is] defined as the ratio in decibels between the acoustic intensity $I(r,z)$ at a field point and the intensity I_0 at 1-m distance from the source,'' uses the MSP definition exclusively for subsequent calculations (p. 84 ff).

III. ALTERNATIVE SONAR EQUATIONS

The missing factor in the sonar equation (Sec. II A) and the ratio between the two pl forms (Sec. II B) are identical—both equal to $Z(\mathbf{x})/Z(\mathbf{x}_0)$. Consequently, if $\mathrm{pl}_{\mathrm{MSP}}$ is used in Eq. (4) instead of $\mathrm{pl}_{\mathrm{EPWI}}$, there is no longer a need for a correction term. However, doing so results in a confusing hybrid sonar equation with some terms defined as EPWI and others as MSP. Two alternative sonar equations are explored below, each using either all EPWI terms (Sec. III A) or all MSP (Sec. III B).

A. EPWI sonar equation

Hall[8] points out that the EPWI source level of Eq. (1) is equal to the source power per unit solid angle (divided by a reference power of $I_{\mathrm{ref}} r_{\mathrm{ref}}^2$). It is convenient to adopt this as a definition of sl, but omitting the reference power in the denominator. In other words, $\mathrm{sl}_{\mathrm{EPWI}} \equiv W/(4\pi)$, for a source power W. (Strictly speaking, the radiated power, for a fixed monopole strength, is a function of the medium in which it is

TABLE I. A summary of definitions and units for the traditional, EPWI, and MSP sonar equations.

Sonar equation term	Traditional sonar equation		EPWI sonar equation		MSP sonar equation	
	Definition	Unit	Definition	Unit	Definition	Unit
sl	$\frac{I_0}{I_{\text{ref}}(\mathbf{x}_0)}$	⋯	$\frac{W}{4\pi}$	pW sr^{-1}	$p_0^2 r_{\text{ref}}^2$	μPa^2 m^2
pl	$\frac{I_0}{I_S^{\text{eq}}(\mathbf{x})}$	⋯	$\frac{W}{4\pi I_S^{\text{eq}}(\mathbf{x})}$	m^2 sr^{-1}	$\frac{p_0^2 r_{\text{ref}}^2}{\overline{p_S(\mathbf{x})^2}}$	m^2
nl	$\frac{I_N^{\text{eq}}(\mathbf{x})}{I_{\text{ref}}(\mathbf{x})}$	⋯	$I_N^{\text{eq}}(\mathbf{x})$	pW m^{-2}	$\overline{p_N(\mathbf{x})^2}$	μPa^2
snr	$\frac{I_S^{\text{eq}}(\mathbf{x})}{I_N^{\text{eq}}(\mathbf{x})}$	⋯	$\frac{I_S^{\text{eq}}(\mathbf{x})}{I_N^{\text{eq}}(\mathbf{x})}$	⋯	$\frac{\overline{p_S(\mathbf{x})^2}}{\overline{p_N(\mathbf{x})^2}}$	⋯

placed. Here W is the power radiated by an equivalent point source of the same monopole strength as the true source, placed in an infinite medium of the same impedance as the true medium.) Further, defining pl_{EPWI} as $W/(4\pi I_S^{\text{eq}})$ and nl_{EPWI} as I_N^{eq}, the EPWI sonar equation can be written as

$$\text{snr}_{\text{EPWI}} \equiv \frac{I_S^{\text{eq}}(\mathbf{x})}{I_N^{\text{eq}}(\mathbf{x})} = \frac{\text{sl}_{\text{EPWI}}}{\text{pl}_{\text{EPWI}} \cdot \text{nl}_{\text{EPWI}}}. \tag{7}$$

These definitions are consistent with those of Hall.[8] They are summarized in Table I.

B. MSP sonar equation

An alternative self-consistent formulation, making use of the MSP definition of pl, can be obtained by replacing all occurrences of EPWI in the above equations with MSP. The resulting sonar equation is

$$\text{snr}_{\text{MSP}} \equiv \frac{\overline{p_S(\mathbf{x})^2}}{\overline{p_N(\mathbf{x})^2}} = \frac{\text{sl}_{\text{MSP}}}{\text{pl}_{\text{MSP}} \cdot \text{nl}_{\text{MSP}}}, \tag{8}$$

where the definitions of sl_{MSP} and pl_{MSP} (see Table I) both refer to the mean-square pressure at distance r_{ref} from the source, p_0^2. This quantity can be defined as

$$p_0^2 \equiv Z(\mathbf{x}_0) \frac{W}{4\pi r_{\text{ref}}^2}.$$

An advantage of this approach (over the EPWI equations) is that the numerical adjustments required, to convert to or from the traditional sonar equation, are small. In fact, if expressed in decibels and for a reference distance r_{ref} of 1 m, this option is identical to Ross' sonar equation. The source level and propagation loss definitions are also consistent with those of Refs. 9, 25, and 26.

IV. CONCLUSIONS

(i) The rigorous application of traditional sonar equation definitions results in the appearance of a factor not normally included in the sonar equation, equal to $Z(\mathbf{x})/Z(\mathbf{x}_0)$. Although this factor is usually close to unity, its omission in extreme cases may lead to significant error.

(ii) The modern *de facto* definition of pl is consistent with most textbooks published since 1975, including Morfey's dictionary. It is not consistent with the definition from Urick's book.

(iii) Two alternatives to the traditional sonar equation are discussed. These are obtained by defining the individual sonar equation terms in such a way as to eliminate the need for the impedance ratio. One of them, based on Hall's definitions, is consistent with the early propagation loss definition. The other uses the modern definition and is equivalent to Ross' sonar equation.

ACKNOWLEDGMENTS

The ideas presented here were stimulated first by Hall's 1995 article and later by numerous discussions with Professor C.L. Morfey of ISVR (the University of Southampton) during and since the compilation of his dictionary. Comments made by Professor Morfey on an earlier version of this article resulted in significant improvements to the final manuscript. Thanks are also due to Joyce Weston for making available important early documents.

[1] R. J. Urick, *Principles of Underwater Sound*, 3rd ed. (Peninsula, Los Altos, 1983).
[2] J. W. Horton, *Fundamentals of SONAR*, 2nd ed. (United States Naval Institute, Annapolis, 1959).
[3] National Defense Research Committee (1946) Physics of Sound in the Sea, Summary Technical Report of Division 6 NDRC, Vol. 8, Washington DC. Reprinted in 1969 by Department of the Navy Headquarters Naval Material Command, Washington DC 20360 (NAVMAT P-9675).
[4] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, Woodbury, NY, 1994).
[5] D. Ross, *Mechanics of Underwater Noise* (Pergamon, New York, 1976).
[6] W. M. Carey, "Standard definitions for sound levels in the ocean," IEEE J. Ocean. Eng. **20**, 109–113 (1995).
[7] W. J. Marshall, "Descriptors of impulsive signal levels commonly used in underwater acoustics," IEEE J. Ocean. Eng. **21**, 108–110 (1996).
[8] M. V. Hall, "Dimensions and units of underwater acoustic parameters," J. Acoust. Soc. Am. **97**, 3887–3889 (1995); erratum: *ibid.* **100**, 673 (1996).
[9] C. L. Morfey, *Dictionary of Acoustics* (Academic, San Diego, 2001).
[10] E. Lamarre and W. K. Melville, "Sound-speed measurements near the ocean surface," J. Acoust. Soc. Am. **96**, 3605–3616 (1994).
[11] R. J. Urick, "Noise signature of an aircraft in level flight over a hydrophone in the sea," J. Acoust. Soc. Am. **52**, 993–999 (1972).
[12] R. W. Young, "Sound pressure in water from a source in air and vice versa," J. Acoust. Soc. Am. **53**, 1708–1716 (1973).

[13] A. A. Hudimac, ''Ray theory solution for the sound intensity in water due to a point source above it,'' J. Acoust. Soc. Am. **29**, 916–917 (1957).
[14] D. I. Paul, ''Acoustical radiation from a point source in the presence of two media,'' J. Acoust. Soc. Am. **29**, 1102–1109 (1957).
[15] M. S. Weinstein and A. G. Henney, ''Wave solution for air-to-water sound transmission,'' J. Acoust. Soc. Am. **37**, 899–901 (1965).
[16] H. Medwin and J. D. Hagy, Jr., ''Helmholtz–Kirchhoff theory for sound transmission through a statistically rough plane interface between dissimilar fluids,'' J. Acoust. Soc. Am. **51**, 1083–1090 (1972).
[17] H. Medwin, R. A. Helbig, and J. D. Hagy, Jr., ''Spectral characteristics of sound transmission through the rough sea surface,'' J. Acoust. Soc. Am. **54**, 99–109 (1973).
[18] D. J. Bordelon, ''Sound pressure in air from a source in water and vice versa,'' J. Acoust. Soc. Am. **55**, 869–870 (1974).
[19] F. R. DiNapoli and R. L. Deavenport, ''Theoretical and numerical Green's function field solution in a plane multilayered medium,'' J. Acoust. Soc. Am. **67**, 92–105 (1980).
[20] F. B. Jensen and W. A. Kuperman, ''Sound propagation in a wedge-shaped ocean with a penetrable bottom,'' J. Acoust. Soc. Am. **67**, 1564–1566 (1980).
[21] R. B. Evans, ''A coupled mode solution for acoustic propagation in a waveguide with stepwise depth variations of a penetrable bottom,'' J. Acoust. Soc. Am. **74**, 188–195 (1983).
[22] A. D. Pierce, ''Augmented adiabatic mode theory for upslope propagation from a point source in variable-depth shallow water overlying a fluid bottom,'' J. Acoust. Soc. Am. **74**, 1837–1847 (1983).
[23] M. A. Pedersen and D. F. Gordon, ''Normal-mode theory applied to short-range propagation in an underwater acoustic surface duct,'' J. Acoust. Soc. Am. **37**, 105–118 (1965).
[24] M. A. Pedersen and D. F. Gordon, ''Normal-mode theory and ray theory applied to underwater acoustic conditions of extreme downward refraction,'' J. Acoust. Soc. Am. **51**, 323–368 (1972).
[25] C. S. Clay and H. Medwin, *Acoustical Oceanography: Principles & Applications* (Wiley, New York, 1977), p. 93–103.
[26] H. Medwin and C. S. Clay, *Fundamentals of Acoustical Oceanography* (Academic, Boston, 1998), p. 113–115.

Tabu for matched-field source localization and geoacoustic inversion

Zoi-Heleni Michalopoulou and Urmi Ghosh-Dastidar
Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, New Jersey 07102

(Received 25 June 2003; revised 29 October 2003; accepted 31 October 2003)

Tabu is a global optimization technique that has been very successful in operations research. In this paper, a Tabu-based method is developed for source localization and geoacoustic inversion with underwater sound data; the method relies on memory to guide the multiparameter search. Tabu is evaluated through a comparison to simulating annealing. Both methods are tested by inverting synthetic data for various numbers of unknown parameters. Tabu is found to be superior to the simulated annealing variant implemented here in terms both of accuracy and efficiency. Inversion results from the SWellEX-96 data set are also presented. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1635408]

PACS numbers: 43.30.Wi, 43.60.Kx [AIT] Pages: 135–145

I. INTRODUCTION

To locate a sound source in the ocean, matched-field processing (MFP)[1,2] is often employed. Correlations are computed between the sound field and replica fields generated for candidate locations. Location estimates are identified by maximizing the correlation.

In the same manner, MFP can be employed to estimate parameters other than location that influence underwater sound propagation. In practice, since the propagation medium is usually only approximately known, MFP can be used for the simultaneous estimation of source location and environmental parameters. The estimation procedure becomes multidimensional, and exhaustive implementation is computationally inefficient. Global optimization methods are employed in order to efficiently search the objective function (measure of correlation between data and replicas). Simulated annealing and genetic algorithms, introduced in matched-field inversion in Ref. 3 and Ref. 4, respectively, are two popular global optimization approaches in the matched-field literature; both methods in several variants have been successfully applied to synthetic and real data problems; simulated annealing results have been presented in Refs. 5–9, and inversion with genetic algorithms has been discussed in Refs. 10–16. Approaches such as neural networks[17,18] and the simplex algorithm[19] have also been tested on MFP. Optimization of posterior probability distribution calculations with Monte Carlo integration[20] or Gibbs sampling[21] have been proposed as well. Hybrid methods combining local and global searches have been discussed in Refs. 22–24, and a statistical freeze-bath approach was presented in Ref. 25.

In this work, we conduct a matched-field search using *Tabu*, an optimization method that has been used extensively and successfully on operations research problems[26] and also on seismic inversion.[27] To our knowledge, MFP results using Tabu have not previously been presented to the underwater acoustics community.

Tabu is a search approach which uses memory in order to navigate the search space efficiently, avoiding areas already visited, especially those where the objective function takes on small values (for maximization). Tabu is a very flexible methodology; the available literature does not dictate how memory should be managed but rather suggests possible memory uses. In this paper, we present our implementation of Tabu optimization and its application to MFP.

Section II presents a description of the Tabu algorithm developed for matched-field inversion. Section III presents results from Tabu optimization on MFP with synthetic data; results are evaluated through a comparison to results obtained using simulated annealing. Section IV presents results from matched field inversion applied to SWellEX-96 data. Conclusions are presented in Sec. V.

II. THE TABU OPTIMIZATION METHOD

A Tabu search starts from a set of initial conditions (in our case, randomly selected values of the unknown parameters within *a priori* chosen bounds) on which the objective function is calculated. Subsequently, a neighborhood is formed around the initial conditions; the neighborhood consists of points obtained from the initial set by perturbing one parameter at a time. The objective function is then evaluated for each neighbor and the process moves to the neighbor with the highest objective function value. The process is a ''steepest ascent–mildest descent'' approach,[28] that is, the neighbor with the highest value of the objective function is selected after the objective function at all neighbors has been calculated, leading to the largest improvement or the smallest degradation possible. A neighborhood is then formed around the selected neighbor; the process continues in the same manner, forming a new neighborhood around the chosen model.

So far, the described maximization approach sounds like a local search. In a multidimensional search space governed by numerous local maxima, this process will most certainly get trapped at a local peak. In order to avoid that, Tabu uses memory, determining which neighbors would be acceptable or prohibited (*tabu* or *taboo*) from consideration. Memory is used through the generation of tabu lists, consisting of tabu moves.

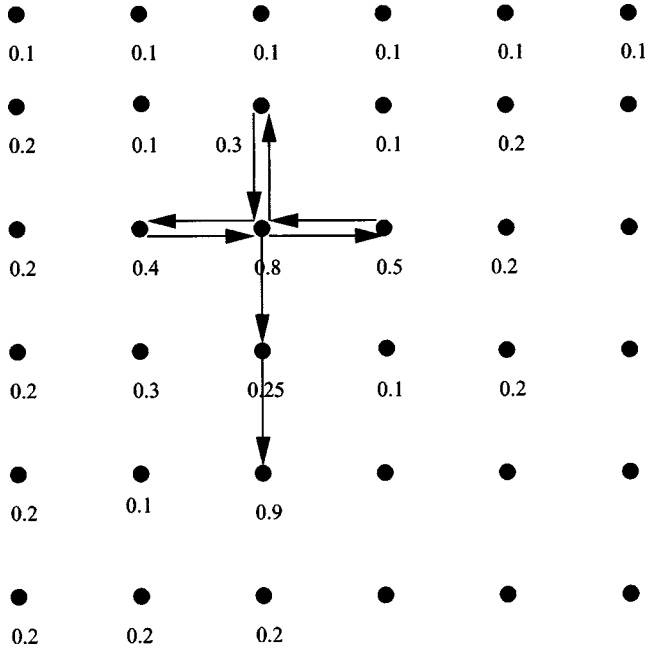


FIG. 1. An illustration of how two-dimensional Tabu proceeds in the search space.

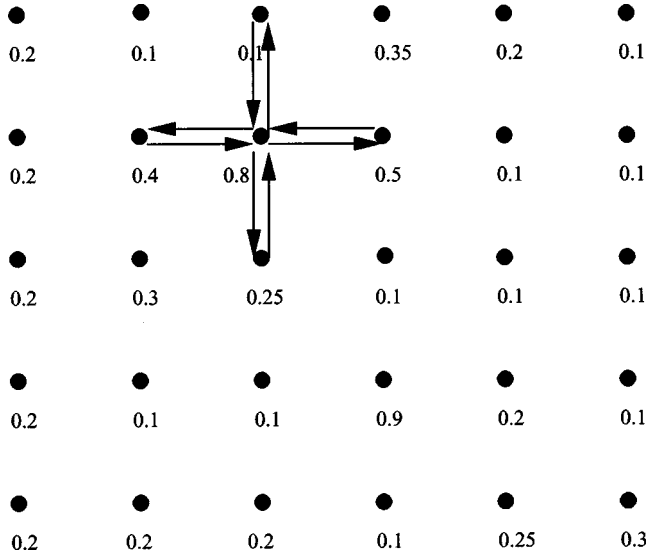


FIG. 2. Two-dimensional Tabu gets trapped in a local maximum.

There is no strict set of guidelines on how lists are generated. There are certain categories of moves that one might want to avoid; for example, if a move is ''good,'' that is, it leads from a lower value of the objective function to a higher one, then one might not want to ''undo'' this move. This is the process suggested in Ref. 27; a tabu list is formed containing N past moves that lead to improvements. These moves are not reversed.

It might also be a good idea to prohibit more moves in addition to reversals of previously made good moves. For example, one might want to avoid repeating moves that have been performed in the recent past, in order to avoid cycling (or trapping) in the area of a local maximum.

In our work we use three tabu lists for memory purposes. One list contains the past N moves that led to an improvement; these moves cannot be reversed. The list size N is $2\times n$, where n is the dimension of the search space, ensuring that all routes that would allow the process to escape a local maximum are examined. The second list contains M past moves, where M is the length of the list; these moves cannot be repeated. We refer to the former as the reverse list and the latter as the forward list. There is a third list that we use, containing sets of parameter values that have been visited in the near past and the corresponding objective function values (matched-field correlation). This list serves as a look-up table; if parameter sets in the list are revisited, the objective function is not recalculated but is taken from the list instead. The list sizes for the forward and look-up lists are chosen in an *ad hoc* manner.

At the end of every iteration the lists are updated; the most recent moves are placed at the top of the list, the oldest moves are ''flushed out'' so that list size is maintained.

The use of memory assists the process in navigating the search space without easily getting trapped in local maxima. Figure 1 shows how the process could escape a local maximum in a two-dimensional space. Using the method outlined above, the method examines neighbors and moves to the neighbor with the highest objective function value, provided that this neighbor is not ''tabu.'' In the neighborhood there are four points, resulting from perturbing one parameter at a time (for an n-dimensional parameter perturbing one parameter at a time would yield $2\times n$ neighbors). Assuming that the two dimensions here correspond to source range r and source depth z_s, the four neighbors would then be $(r+\delta r, z_s)$, $(r-\delta r, z_s)$, $(r, z_s+\delta z_s)$, and $(r, z_s-\delta z_s)$.

We set the search to start from a set of values for r and z_s for which the objective function is 0.8. The process moves then to the neighbor with the 0.5 value of the objective function. From there, the process moves back to the original 0.8 value, since it is the neighbor with the highest value. The procedure cannot revisit 0.5 after that, because the move from 0.5 to 0.8 was a ''good'' move (it led to an improvement); it is thus stored in a tabu list and cannot be reversed. Since going from 0.8 to 0.5 is now tabu, the process picks the next best neighbor, which has an objective function value of 0.4. From 0.4 the process moves back to 0.8. Moving back to point 0.4 is now tabu, so the process moves to 0.3 and then back to 0.8. The next non-tabu place is the one at 0.25. The process moves there and can subsequently escape to find a higher value, that of 0.9. In this way, the optimization process escapes the local maximum of 0.8 and explores more areas of the search space.

Although in the particular example the process is successful, it is often the case that a trapping will occur despite memory use and intelligent navigation. Figure 2 demonstrates a case where a local maximum is ''isolated'' with the search process not being able to escape. In order to avoid such problems, we propose that the process detects such traps. When a trap is identified, the process continues the search using a different (larger) neighborhood width; such moves are termed ''jumps.'' We have found that using random jumps assists the process to navigate more efficiently in the search space (the use of jumps of a deterministically selected size could lead to more trappings). Figure 3 illus-

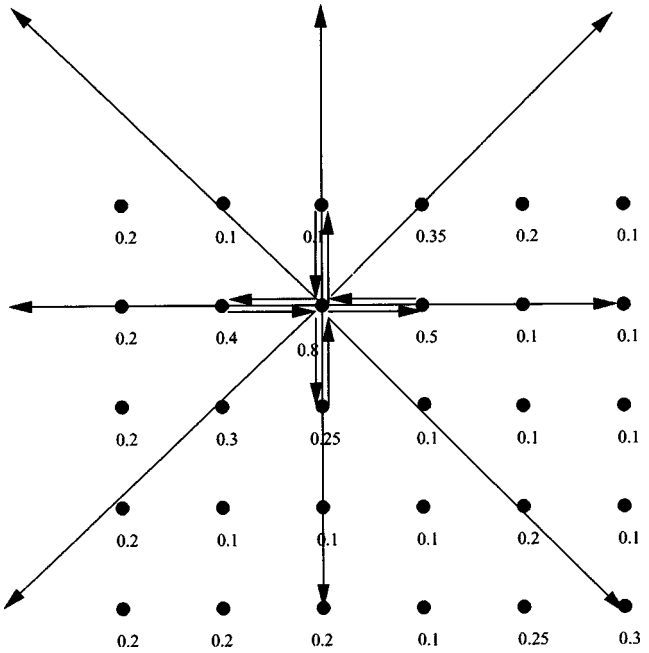


FIG. 3. Dynamic neighborhood change to avoid traps in local maxima.

trates such jumps. As can be seen in the figure, in case of a jump, the new broader neighborhood contains ''diagonal'' transitions as well. That is, when the neighborhood width changes because a trap has been detected, we are not restricted into changing one parameter at a time as we do in Fig. 1; our neighborhoods also include simultaneous changes of all parameters. For the two-dimensional problem that we have illustrated in Figs. 1, 2, and 3, the neighborhood during a jump will consist of eight neighbors (for an n-dimensional search the number of neighbors at a jump would be $2\times n+2^n$). The particular neighbors will be $(r+k\delta r, z_s)$, $(r-k\delta r, z_s)$, $(r, z_s+k\delta z_s)$, $(r, z_s-k\delta z_s)$, $(r+k\delta r, z_s+k\delta z_s)$, $(r+k\delta r, z_s-k\delta z_s)$, $(r-\delta r, z_s+k\delta z_s)$, $(r-k\delta r, z_s-k\delta z_s)$. Scalar quantity k is a random number larger than 1. Additional experimentation with Tabu revealed to us that we can accelerate the identification of the global maximum, if, every time we jump, we further increase the neighborhood size by examining several (three, in our case) neighborhoods simultaneously. For the two-dimensional problem, the first neighborhood consists of the eight neighbors we just listed. We generate the two additional neighborhoods (consisting of eight neighbors each) by doubling and tripling the scalar k. Thus, for an n-dimensional search the number of neighbors examined during a jump would be $3\times(2\times n+2^n)$.

Figure 2 also illustrates the choice of $2\times n$ for the length of the list containing the improvements. A point in the search space has $2\times n$ neighbors. If this point is a local maximum, there are $2\times n$ moves (from the $2\times n$ neighbors to that point) that would lead to an improvement. Thus, $2\times n$ is the size of the improvement list, so that all paths around a local maximum are investigated.

Figure 4 shows a two-dimensional Bartlett ambiguity surface generated with synthetic data. Superimposed on the surface are those source range and depth points visited by a single Tabu run during the search; the search started from an initial range value of 4.98 km and an initial source depth value of 2 m. As the color bar shows, dark areas correspond to high-ambiguity surface values. As Tabu makes selections in the space, we can see from the figure that the visited areas are high-ambiguity surface value areas. For the demonstrated case, the true source location is at 2 km in range and 100 m in depth, values that are identified by Tabu as can be seen from the numerous circles in that region in Fig. 4.

In matched-field inversion, it is often the case that a search encounters ridges that are obliquely positioned with respect to the search axes. The direction of the ridges is determined by correlations between the unknown parameters. Figure 5 demonstrates how Tabu performs in such areas, quickly identifying the direction of the ridge to find the maximum. The figure shows a detail of the ambiguity surface of Fig. 4 in the vicinity of the true global maximum. Tabu starts from an initial condition of 1.92 km in range and 92 m

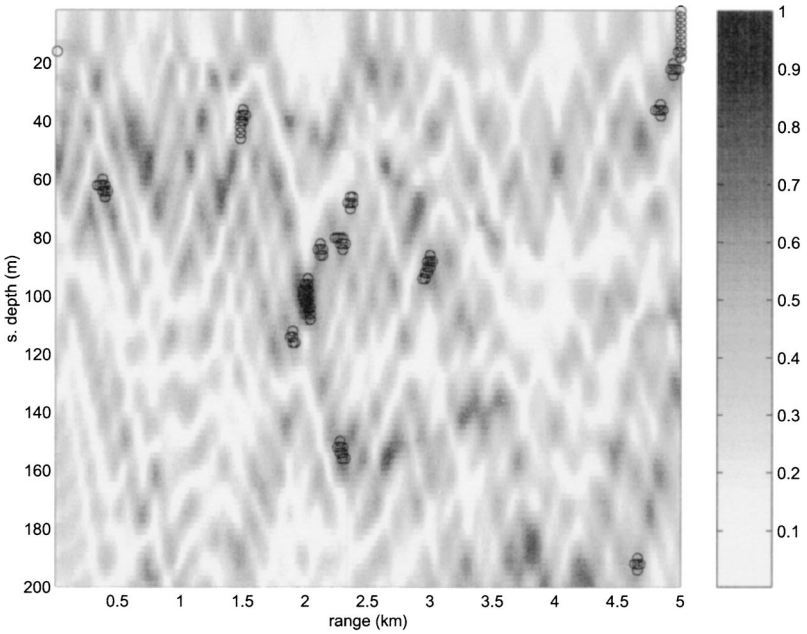


FIG. 4. A range-depth ambiguity surface with the superimposed points (indicated by circles) on the range–depth space visited by Tabu.

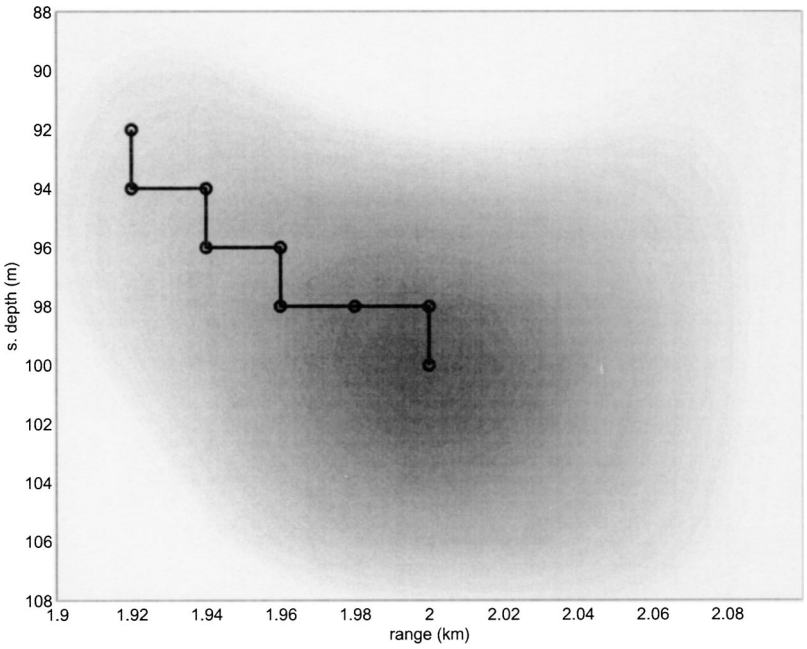


FIG. 5. A detail of the range–depth ambiguity surface showing how Tabu climbs ridges obliquely positioned with respect to the search axes because of parameter correlations; circles indicate the points visited by Tabu.

in source depth and climbs up to the maximum located at 2 km and 100 m in range and depth, respectively.

III. TABU FOR MATCHED-FIELD PROCESSING USING SYNTHETIC DATA

To evaluate Tabu, we generated noise-free, acoustic field data using KRAKEN.[29] The data were generated for a frequency of 150 Hz, a source range and depth of 2 km and 100 m, and an ocean depth of 216.5 m. The first receiver was at a depth of 50 m; all phones were vertically separated with a spacing of 10 m. The environment was that of SWellEX-96.[30–33] Considering no noise, there was full control of the value (equal to 1) and location of the global maximum of the ambiguity surface. A sketch of the environment is shown in Fig. 6.

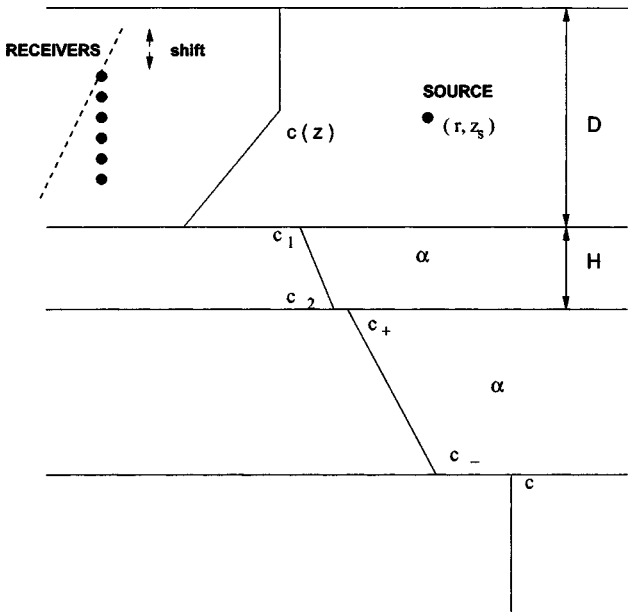


FIG. 6. The environment used in the simulations. The environment is similar to that of the SWellEX-96 experiment.

The true values, intervals, and increments for all parameters considered here are presented in Table I. The increments were selected empirically after studying the sensitivity of the matched-field correlation to the parameters of interest. The spacings were chosen so that the maximization process would not miss important peaks because of undersampling.

The Bartlett processor[2] was used for the computation of the objective function f

$$f(\mathbf{q}) = \mathbf{w}(\mathbf{q})^* \mathbf{d}\mathbf{d}^* \mathbf{w}(\mathbf{q}), \tag{1}$$

where $\mathbf{q}$ is the vector of parameters to be estimated, $\mathbf{d}$ is a data vector received at a set of hydrophones, and $\mathbf{w}$ is a normalized replica.

For comparison to Tabu, the efficient simulated annealing method of Ref. 3 was implemented. The cooling rate was linear. Simulated annealing in any variant requires the selection of a ''suitable'' initial temperature and parameter perturbation sizes. All these are, typically, empirically determined. To ensure a fair comparison, we first experimented with simulated annealing to see how many iterations (and forward models) were required by the method to identify high

TABLE I. True values, search intervals, and increments for unknown parameters.

Parameter	True value	Interval	Increment
r (km)	2	0–5	0.02
z_s (m)	100	0–200	2
D (m)	216.5	200–230	2
Tilt (deg)	0	−5–5	0.5
Sed. (m)	23.5	5–150	5
Shift (m)	0	−10–10	2
Att. (dB/mkHz)	0.2	0.01–2	0.1
c_1 (m/s)	1572	1530–1630	5
c_2 (m/s)	1593	1550–1650	5

TABLE II. Simulated annealing and Tabu results for the 3D test case.

SA					TABU				
r	z_s	D	Max	Models	**r**	**z_s**	**D**	**Max**	**Models**
2.00	100.2	216.5	1.00	2921	**1.96**	**99.9**	**214.6**	**0.97**	**856**
1.94	99.4	215.3	0.80	367	**2.00**	**99.9**	**216.7**	**1.00**	**733**
2.02	100.4	217.6	1.00	1525	**1.93**	**98.0**	**212.5**	**0.96**	**450**
2.01	99.8	216.4	0.98	2172	**2.06**	**102.4**	**219.8**	**0.98**	**1379**
1.93	98.1	212.5	0.93	773	**1.99**	**99.3**	**216.6**	**0.99**	**1852**
1.99	101.3	215.9	0.94	1206	**2.05**	**100.9**	**219.2**	**0.98**	**365**
2.03	102.6	218.8	0.94	1380	**2.05**	**100.8**	**219.0**	**0.99**	**303**
2.16	105.9	225.4	0.75	1731	**1.96**	**98.8**	**214.9**	**0.98**	**113**
2.03	101.0	218.1	0.99	2423	**1.95**	**98.3**	**213.7**	**0.97**	**204**
2.01	99.7	217.4	0.99	2228	**2.11**	**103.8**	**222.7**	**0.93**	**1788**

matched-field correlation areas, where the number of forward models is defined as the number of normal-mode calculations. The goal was to implement a simulated annealing process that would have the opportunity to explore a large part of the search space, without becoming an exhaustive search approach. Initial temperatures were selected that allowed the process to perform several downhill moves, in order for it to be able to escape local maxima and have a good chance of being directed towards the true values of the unknown parameters. The random parameter perturbations were cubed to support mostly small and less frequently larger parameter variations.[3] Once a reasonable number of forward model calculations for simulated annealing was identified, Tabu was run for the same number of forward model calculations. The simulated annealing estimates considered as the output of the simulated annealing approach were those values that gave the maximum matched-field correlation during the search process; most times, these values coincided with the values at which convergence was achieved. Sometimes, however, simulated annealing would visit high correlation areas at the early stage of the annealing process, when the temperature was high; subsequently, it would move into other regions of the search space and would converge to areas corresponding to a lower matched-field correlation. The Tabu estimates were those values visited by Tabu yielding the highest correlation within the predetermined number of models.

Initially, a three-dimensional search was performed for source range, source depth, and water column depth with both Tabu and simulated annealing. Table II presents the results obtained from ten runs (corresponding to ten sets of initial conditions) for simulated annealing and Tabu (in boldface) applied to the maximization of the Bartlett ambiguity surface. Each run is started from a different set of initial conditions (however, exactly the same sets of initial conditions were used for Tabu and simulated annealing). The table presents estimates for the three unknown parameters, the maximum objective function achieved for each run, and the number of models necessary for the reported maximum to be obtained. The maximum number of models considered was 3000.

In all cases, the estimates of the unknown parameters are close to the true parameter values for both methods. In two cases, however, a relatively low value of Bartlett correlation is reached by simulated annealing. Comparing the results from the two approaches, two observations can be made: (i) Tabu requires fewer objective function calculations than simulated annealing to identify the parameter values at which the maximum occurs, and (ii) Tabu reaches deeper

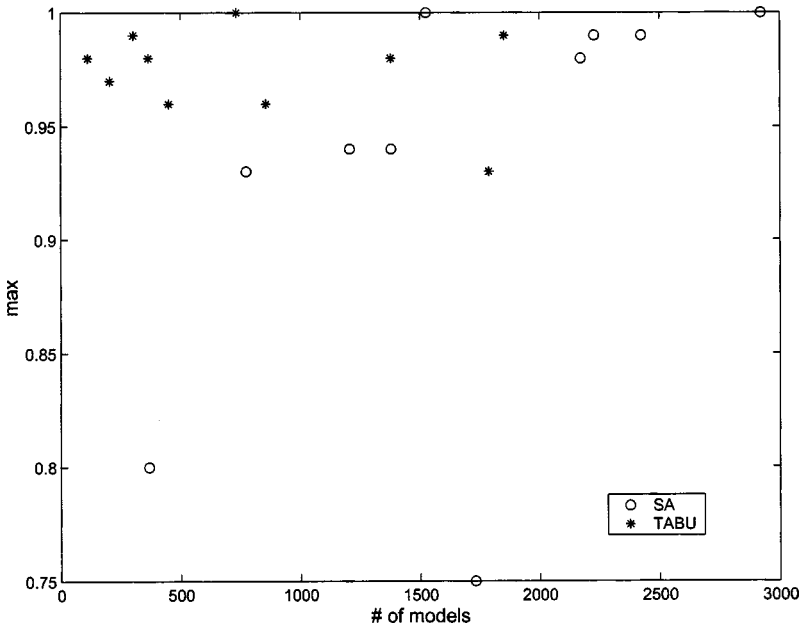


FIG. 7. Tabu-simulated annealing comparison for the three-dimensional case: Achieved maximum vs required number of models.

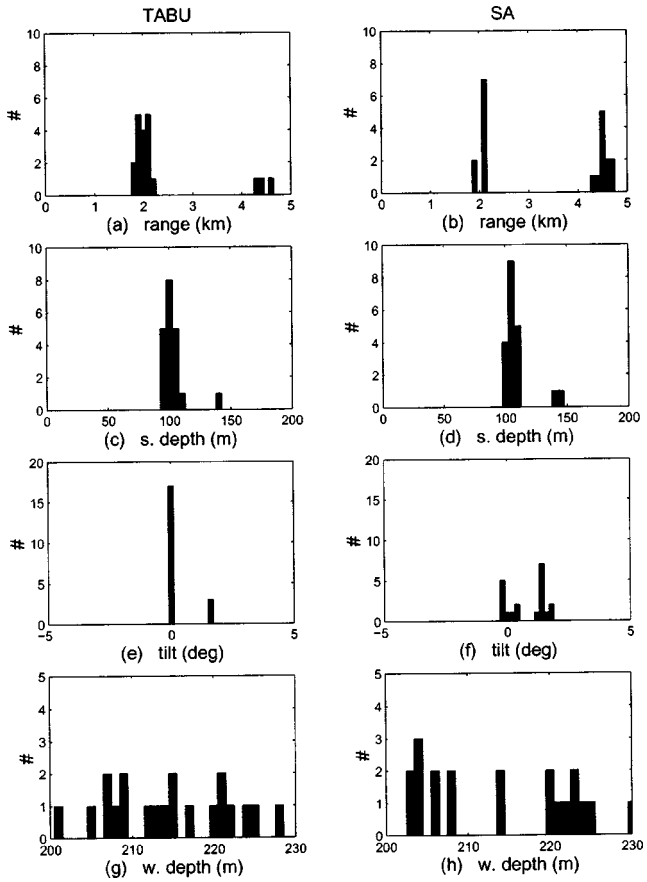


FIG. 8. Tabu-simulated annealing comparison for a search in seven dimensions [(a)–(b) source range; (c)–(d) source depth; (e)–(f) array tilt, and (g)–(h) water depth].

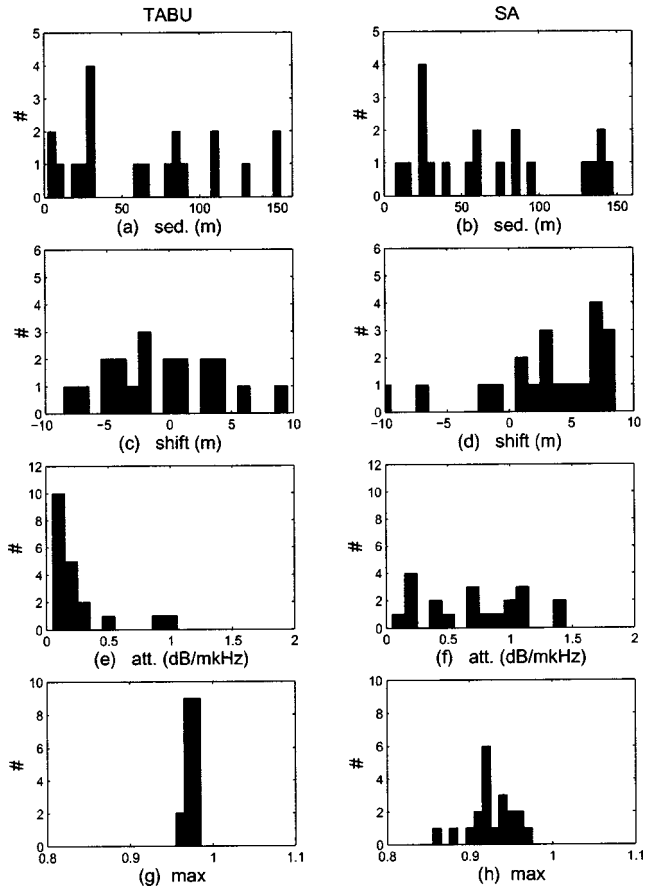


FIG. 9. Tabu-simulated annealing comparison for a search in seven dimensions [(a)–(b) sediment thickness; (c)–(d) receiver shift; (e)–(f) attenuation in sediment; and (g)–(h) achieved maximum].

into the objective function than simulated annealing, that is, it performs an effective local search once the general area of the global maximum is identified, obtaining high objective function values in the identified area. This is also shown in Fig. 7, which shows the attained maxima vs the numbers of models necessary to reach these maxima for the two methods.

For the next test, inversion was performed in a larger search space; four additional parameters (receiving array tilt, vertical receiver shift, first sediment thickness, and compressional attenuation) were considered unknown. The methods were evaluated for a maximum number of 50 000 models.

For the seven-dimensional case, according to Sec. II, when the process needed to ''jump'' because of identified local maxima, the number of the neighborhood points would be $3\times2\times7+3\times2^7$; the $3\times2\times7$ neighbors correspond to the one-at-a-time parameter perturbations, whereas the 3×2^7 neighbors correspond to the ''diagonal'' perturbations which involve perturbing all parameters simultaneously. The size of the neighborhood becomes large, delaying the optimization process. Observation of the effect of parameter changes on the field and objective function allowed a reduction in the neighborhood size at a jump. The field was more sensitive to parameters such as source range, depth, tilt, and water column depth than to sediment thickness, phone shift, and attenuation. For the seven (and higher)-dimensional searches, we included in our neighborhoods (when a jump needs to be implemented) single parameter perturbations and also simultaneous parameter perturbations of only the four most significant parameters. The neighborhood size became $3\times(2\times7+2^4)$.

Figures 8 and 9 present histograms from the seven-dimensional estimates. Source range and depth, tilt, and water column depth are shown in Fig. 8; sediment thickness, receiver shift, and attenuation are shown in Fig. 9. Histograms of the achieved maxima for the 20 cases for the two methods are also shown in Fig. 9 (bottom row). From the range, source depth, and tilt estimates for Tabu [Figs. 8(a), (c), and (e)], one can see that Tabu mostly pinpoints quite accurately these three parameters. Only in three cases does Tabu get trapped in local maxima at roughly 4.5 km in source range, between 100 and 150 m in source depth, and 1.5 degrees in tilt. These are substantial local maxima, comparable in terms of correlation to the optimal value of 1; as can be seen in Fig. 9(g), all maxima identified by Tabu are between 0.96 and 1.

Simulated annealing, on the other hand, identifies the true global maximum only half of the time; for the other half, it gets trapped in the same local maxima as Tabu did three times. The correlations obtained by simulated annealing are also quite high [Fig. 9(h)] but lower than those obtained with Tabu. Figure 10 shows the maxima obtained vs the necessary number of forward model calculations for the two methods, again demonstrating that Tabu achieves, in general, higher

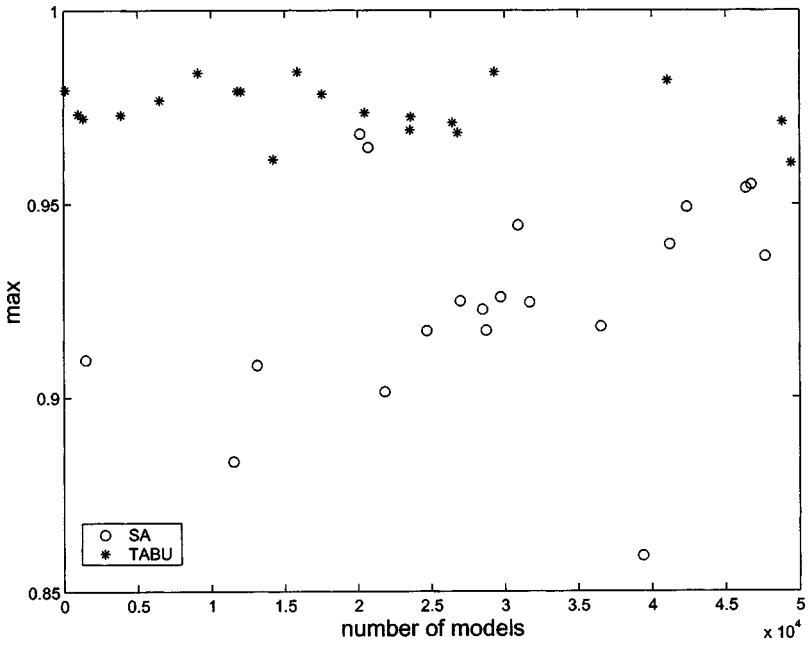


FIG. 10. Tabu-simulated annealing comparison for the seven-dimensional case: Achieved maximum vs required number of models.

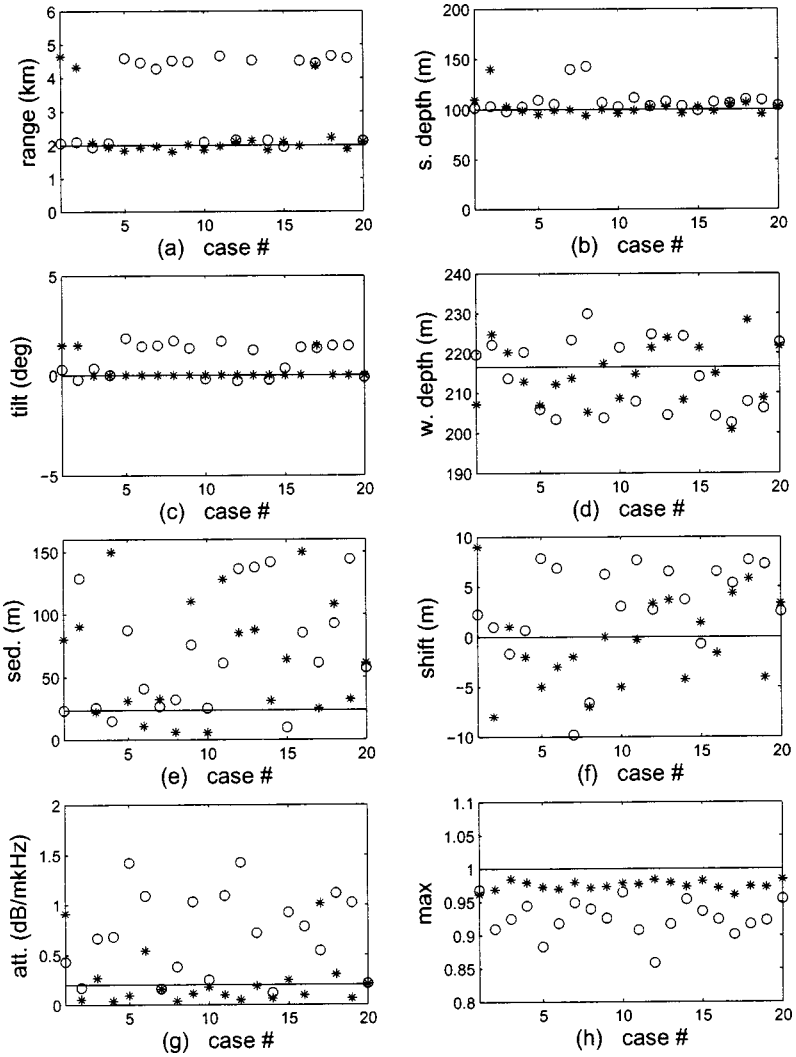


FIG. 11. Tabu-simulated annealing comparison for the seven-dimensional case: Solid lines show the true values of the parameters; circles show the simulated annealing estimates, and asterisks show the TABU estimates.

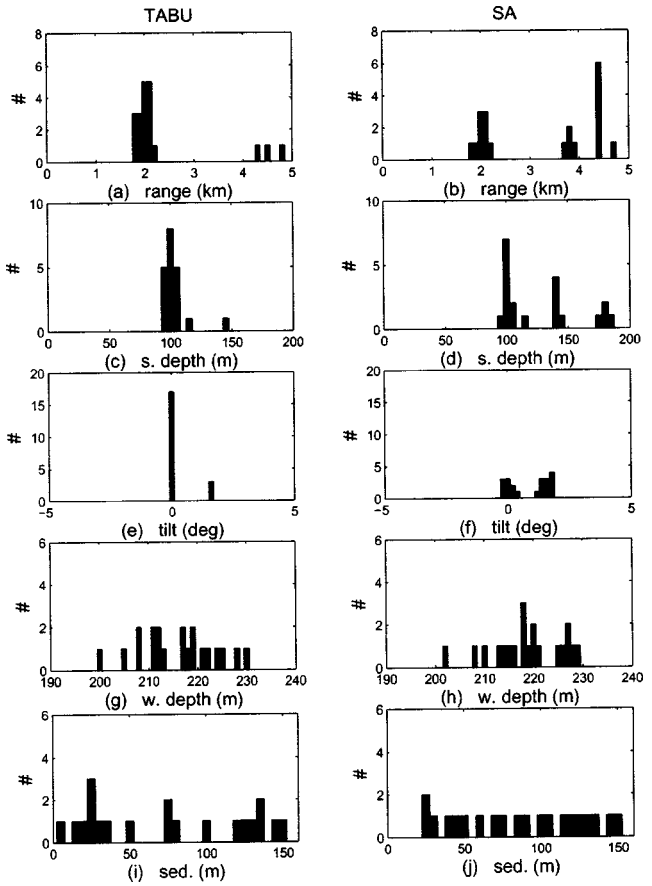


FIG. 12. Tabu-simulated annealing comparison for a search in nine dimensions [(a)–(b) source range; (c)–(d) source depth; (e)–(f) array tilt; (g)–(h) water depth; and (i)–(j) sediment thickness].

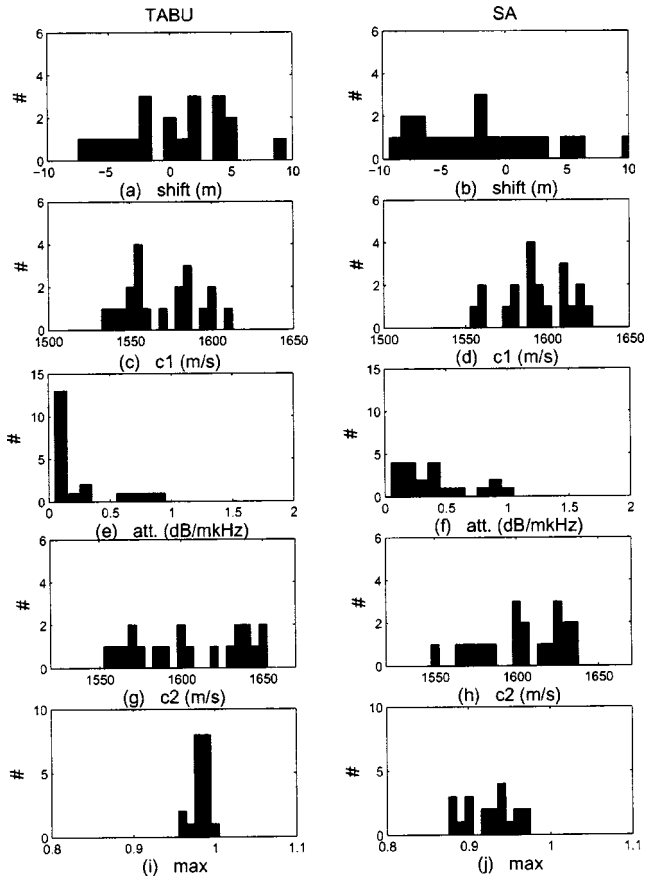


FIG. 13. Tabu-simulated annealing comparison for a search in nine dimensions [(a)–(b) receiver shift; (c)–(d) sound speed at top of sediment; (e)–(f) attenuation; (g)–(h) sound speed at bottom of sediment; and (i)–(j) achieved maximum].

matched-field correlations than simulated annealing for a smaller number of models. In most cases, Tabu needs fewer than 30 000 models to exceed a correlation of 0.96. Simulated annealing reaches lower correlation levels and requires more forward model computations. Achieving high matched-field correlations is a result of the ability of Tabu to operate successfully on a local level as well as on a global level; its neighborhood structure allows it to examine closely and in detail high correlation areas and further refine the search.

The results of the seven-dimensional inversion are also summarized in Fig. 11. The figure shows the true values of the seven parameters (solid lines) and the Tabu and simulated annealing estimates for the 20 examined cases; the maxima are also shown. As the plots indicate, there is a very good fit between Tabu estimates and true values for source range, source depth, and tilt (with the exception of the three cases previously mentioned); simulated annealing estimates often deviate from the true values. Tabu estimates of other parameters vary and are not as close to the true values as the estimates for source range, depth, and array tilt. The lack of a good fit for those parameters (common to both methods) is attributed to their relatively small impact on the acoustic field. However, the overall attained maximum correlation is higher (and closer to 1) in Tabu than in simulated annealing.

Inversion was performed for nine parameters, adding sound speed at the top and bottom of the first sediment to the seven parameters previously considered. Figures 12 and 13 present histograms obtained from the estimates of Tabu and simulated annealing. The maximum number of models for the evaluation was set to 100 000.

The results are similar to those of the seven-dimensional inversion. Tabu identifies the true values of the parameters in the majority of the runs, reporting 3 out of 20 times local maxima similar to those of the seven-dimensional case (range between 4 and 5 km, source depth 100 and 150 m, and tilt at 1.5 degrees). Simulated annealing identifies the true source location 9 out of 20 times only, whereas it provides estimates associated with the aforementioned local maxima and also with a local maximum at 3.8 km in range, 180 m in source depth, and 1.5 degrees in tilt.

Figure 14 shows the maxima vs models for the two methods, demonstrating that Tabu attains high maxima with fewer forward model calculations than simulated annealing. The relations between estimates and true parameter values are demonstrated in Fig. 15. Tabu estimates are very close to the true values, especially for source range, depth, and array tilt, whereas simulated annealing results vary around the true values, since the method frequently gets trapped in local maxima.

IV. TABU INVERSION WITH SWellEX-96 DATA

Tabu was also applied to data from the SWellEX-96 experiment. The experiment was performed in a shallow water

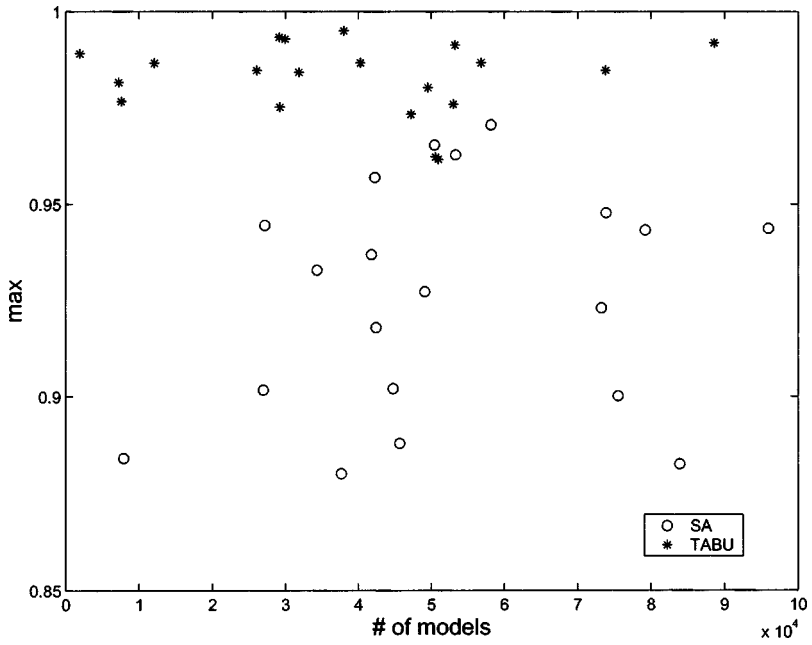


FIG. 14. Tabu-simulated annealing comparison for the nine-dimensional case: Achieved maximum vs required number of models.

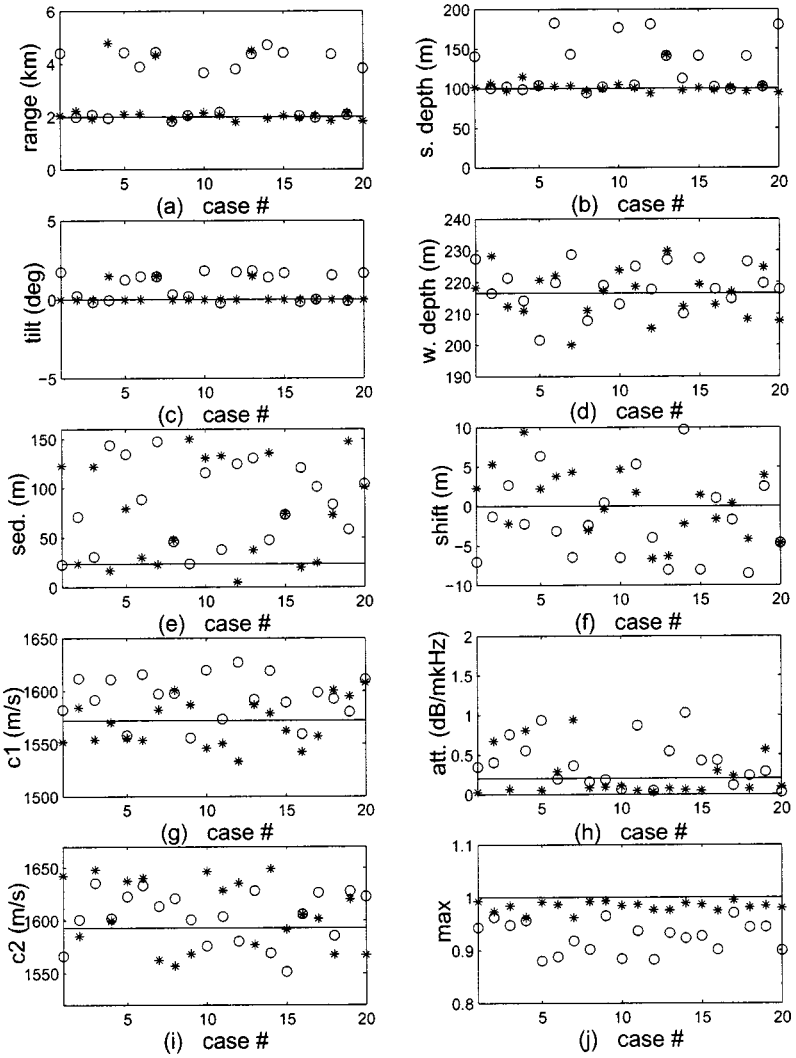


FIG. 15. Tabu-simulated annealing comparison for the nine-dimensional case: Solid lines show the true values of the parameters; circles show the simulated annealing estimates, and asterisks show the TABU estimates for the parameters.

TABLE III. Ground-truth values for SWellEX-96.

r (km)	1.1
z_s (m)	54.6
D (m)	216.5
Sed. (m)	23.5
Att. (dB/mkHz)	2.0
c_1 (m/s)	1572
c_2 (m/s)	1593

(around 216 m deep) environment, which is range independent for the part examined in this work. The environment includes a relatively thin sediment underneath the water column and a thicker second sediment (the properties of which are not investigated here, because it was previously observed that they have negligible impact on the acoustic field[34]). Field measurements were made at vertically spaced hydrophones, with the shallowest phone located near the middle of the water column. Broadband data measurements with frequency content between 200 and 400 Hz were used for the inversion.

Aiming at sidelobe suppression, a broadband Bartlett processor, f_b, was implemented as in Ref. 35

$$f_b(\mathbf{q}) = \frac{\Sigma_f \log_{10}(\mathbf{w}_f^* \mathbf{d}_f \mathbf{d}_f^* \mathbf{w}_f)}{N_f}, \tag{2}$$

where N_f is the number of frequencies, $\mathbf{w}_f$ is the replica at frequency f, and $\mathbf{d}_f$ are the data at frequency f.

Prior knowledge sets the source at around 1 km from the receiving array; the source depth is approximately 55 m. MFP results presented in Ref. 34 are in agreement with the prior beliefs on source location.[30] (However, the matched-field inversion in Ref. 34 did not consider array tilt.) The ground-truth information for the experiment is summarized in Table III. Environmental information was obtained from a geophysical database compiled for the site;[33] source location was estimated during the experiment.[30]

Before applying Tabu to the data, a four-dimensional exhaustive search was first performed for the most important parameters for the search (source range, source depth, tilt, and water column depth). This search resulted in a maximum of -0.61 for a source range of 1.31 km, a source depth of 71 m, a receiving array tilt of 2 degrees, and a water column depth of 218.5 m. Environmental parameters were set to values believed to be descriptive of the particular environment (as was observed in Ref. 34, the most important parameters affecting the acoustic field were source location parameters and water column depth). It was observed that considering tilt as an unknown caused a ''shift'' in the range and source depth estimates; a three-dimensional search (excluding tilt) resulted in values of 1.05 km and 54 m, respectively, for source range and depth. However, the four-dimensional search produced a higher broadband Bartlett value (-0.61) than that obtained for the three-dimensional search (-0.79).

Table IV shows the results obtained for four-, seven-, and nine-dimensional inversion along with the broadband Bartlett correlation values and the numbers of mode calculations necessary for the inversion. The source location estimates and maxima achieved are consistent with the exhaustive four-dimensional search results. The bottom depth is well estimated (see Table III). The sediment thickness and attenuation estimates vary between the seven- and nine-dimensional searches, and also differ from the ground-truth information on these parameters; this is attributed to the small impact these parameters have on the field. The estimate for the sound speed at the top of the sediment (1578 m/s) is very close to the expected values for that parameter (1572 m/s). The sound-speed estimate at the bottom of the sediment varies from what is expected (estimate is 1640 m/s and expected value is 1593 m/s), but is a physically reasonable value.

TABLE IV. Tabu results for SWellEX-96.

	4D	7D	9D
r (km)	1.32	1.31	1.31
z_s (m)	72.8	72.5	72.5
D (m)	219.3	218.0	217.0
Tilt (deg)	1.6	1.5	1.9
Shift (m)	N/A	-1.0	-2.0
Sed. (m)	N/A	10.0	119.9
Att. (dB/mkHz)	N/A	0.7	0.1
c_1 (m/s)	N/A	N/A	1578
c_2 (m/s)	N/A	N/A	1640
Max	-0.61	-0.60	-0.59
Models	180	13677	2304

V. CONCLUSIONS

A global optimization method based on Tabu search was designed and implemented for multidimensional matched-field inversion. The process is simple to implement, relying on neighborhood searches that shift on grids inside the parameter space. The use of memory that Tabu exploits is a novel and powerful feature in matched-field inversion. Through two memory lists, the process is guided away from already examined regions and towards high correlations. Thus, the method does not expend resources in areas of the parameter space that will not offer an improvement to the search. A third list contributes to efficiency by storing previously visited models and corresponding correlations to obviate objective function recalculation. A jump mechanism assists the process to escape local maxima.

The method was tested on both synthetic and real data, and in both cases accurate results were obtained. A comparison showed that the new method is more efficient in terms of forward model calculations and more accurate than an efficient simulated annealing variant. The results demonstrate that Tabu is a powerful global optimization tool for successful matched-field inversion.

Several optimization methods are successful on a global level, but they sometimes require a combination with a local method for further refinement of the matched-field results. Tabu does not have this requirement. In addition to successfully identifying the general area of the highest correlation in the search space, Tabu also performed very well on a local level. Because of its neighborhood search examining all points surrounding a model, Tabu was able to penetrate deep in areas of high correlation yielding high matched-field values.

ACKNOWLEDGMENTS

This work was funded by the Office of Naval Research through Grants N00014-01-1-0085 and N00014-97-1-0600.

[1] A. Baggeroer, W. Kuperman, and P. N. Michalevsky, ''An overview of matched field processing in ocean acoustics,'' IEEE J. Ocean. Eng. **18**, No. 4, 401–424 (1993).
[2] A. Tolstoy, *Matched Field Processing for Underwater Acoustics* (World Scientific, Singapore, 1993).
[3] M. D. Collins and W. Kuperman, ''Focalization: Environmental focusing and source localization,'' J. Acoust. Soc. Am. **90**, 1410–1422 (1991).
[4] P. Gerstoft, ''Inversion of seismoacoustic data using genetic algorithms and *a posteriori* probability distributions,'' J. Acoust. Soc. Am. **95**, 770–782 (1994).
[5] M. D. Collins and W. A. Kuperman, ''Nonlinear inversion for ocean bottom properties,'' J. Acoust. Soc. Am. **92**, 2770–2783 (1992).
[6] C. E. Lindsay and N. R. Chapman, ''Matched field inversion for geoacoustic model parameters using adaptive simulated annealing,'' IEEE J. Ocean. Eng. **18**, No. 3, 224–231 (1993).
[7] N. R. Chapman and C. E. Lindsay, ''Matched field inversion for geoacoustic model parameters in shallow water,'' IEEE J. Ocean. Eng. **21**, 347–355 (1996).
[8] M. R. Fallat and S. E. Dosso, ''Geoacoustic inversion for the Workshop 97 benchmark test cases using simulated annealing,'' J. Comput. Acoust. **6**, 29–43 (1998).
[9] T. B. Neilsen, ''An iterative implementation of rotated coordinates for inverse problems,'' J. Acoust. Soc. Am. **113**, 2574–2586 (2003).
[10] D. Gingras and P. Gerstoft, ''Inversion for geometric and geoacoustic parameters in shallow water: Experimental results,'' J. Acoust. Soc. Am. **97**, 3589–3598 (1995).
[11] G. J. Heard, D. Hannay, and S. Carr, ''Genetic algorithm inversion of the 1997 geoacoustic inversion workshop test case data,'' J. Comput. Acoust. **6**, 61–71 (1998).
[12] P. Ratilal, P. Gerstoft, and J. T. Goh, ''Subspace approach to inversion by genetic algorithms involving multiple frequencies,'' J. Comput. Acoust. **6**, 83–115 (1998).
[13] M. Siderius, P. Gerstoft, and P. Nielsen, ''Broadband geoacoustic inversion from sparse data using genetic algorithms,'' J. Comput. Acoust. **6**, 117–134 (1998).
[14] D. G. Simmons and M. Snellen, ''Multifrequency matched field inversion of benchmark data using a genetic algorithm,'' J. Comput. Acoust. **6**, 135–150 (1998).
[15] M. I. Taroudakis and M. G. Markaki, ''Bottom geoacoustic inversion by broadband matched field processing,'' J. Comput. Acoust. **6**, 167–183 (1998).
[16] V. Westerlin, ''Multifrequency inversion of synthetic transmission loss data using a genetic algorithm,'' J. Comput. Acoust. **6**, 205–221 (1998).
[17] Y. Stephan, X. Demoulin, and O. Sarzeaud, ''Neural direct approaches for geoacoustic inversion,'' J. Comput. Acoust. **6**, 151–166 (1998).
[18] J. M. Ozard, P. Zakarauskas, and P. Ko, ''An artificial neural network for range and depth discrimination in matched field processing,'' J. Acoust. Soc. Am. **90**, 2658–2663 (1991).
[19] M. Musil, N. R. Chapman, and M. J. Wilmut, ''Range-dependent matched field inversion of SWellEX-96 data using the downhill simplex algorithm,'' J. Acoust. Soc. Am. **106**, 3270–3281 (1999).
[20] J. A. Shorey and L. W. Nolte, ''Wideband optimal *a posteriori* probability source localization in an uncertain shallow ocean environment,'' J. Acoust. Soc. Am. **103**(1), 355–361 (1998).
[21] S. E. Dosso, ''Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampling approach,'' J. Acoust. Soc. Am. **111**, 129–142 (2002).
[22] P. Gerstoft, ''Inversion of acoustic data using a combination of genetic algorithms and the Gauss–Newton approach,'' J. Acoust. Soc. Am. **97**, 2181–2190 (1995).
[23] M. R. Fallat and S. E. Dosso, ''Geoacoustic inversion via local, global, and hybrid algorithms,'' J. Acoust. Soc. Am. **105**, 3219–3230 (1999).
[24] S. E. Dosso, M. J. Wilmut, and A. S. Lapinski, ''An adaptive hybrid algorithm for geoacoustic inversion,'' IEEE J. Ocean. Eng. **26**, 323–336 (2001).
[25] L. Jaschke and N. R. Chapman, ''Matched field inversion of broadband data using the freeze bath method,'' J. Acoust. Soc. Am. **106**, 1838–1851 (1999).
[26] F. Glover and M. Laguna, TABU Search (Kluwer Academic, Dordrecht, 1997).
[27] R. Vinther and K. Mosegaard, ''Seismic inversion through tabu search,'' Geophys. Prospect. **44**, 555–570 (1996).
[28] P. Hansen, ''The steepest ascent mildest descent heursitic for combinatorial programming,'' in Congress on Numerical Methods in Combinatorial Optimization (Italy), 1996.
[29] M. B. Porter, ''The KRAKEN normal mode program.'' SACLANT Undersea Research Centre Memorandum (SM-245) and Naval Research Laboratory Mem. Report 6920, 1991.
[30] ''SWellEX-96 Preliminary Data Report,'' Technical Report, 1996.
[31] N. O. Booth, P. A. Baxley, J. A. Rice, P. W. Schey, W. S. Hodgkiss, G. L. D. Spain, and J. J. Murray, ''Source localization with broadband matched field processing in shallow water,'' IEEE J. Ocean. Eng. **21**, 402–412 (1996).
[32] P. A. Baxley, ''Matched field geoacoustic parameter inversion using horizontal line arrays,'' in Fourth International Conference on Theoretical and Computational Acoustics (Trieste, Italy), 1999.
[33] R. T. Bachman, P. W. Schey, N. O. Booth, and R. J. Ryan, ''Geoacoustic databases for matched field processing: Preliminary results in shallow water off San Diego, California,'' J. Acoust. Soc. Am. **99**(4), 2077–2085 (1996).
[34] Z.-H. Michalopoulou, ''Matched impulse response processing for shallow water localization and geoacoustic inversion,'' J. Acoust. Soc. Am. **108**(5), 2082–2090 (2000).
[35] A. Baggeroer, W. Kuperman, and H. Schmidt, ''Matched field processing: Source localization in correlated noise as an optimum parameters estimation problem,'' J. Acoust. Soc. Am. **83**, 571–587 (1988).

Propagation of ultrasonic guided waves in lap-shear adhesive joints: Case of incident a_0 Lamb wave

Francesco Lanza di Scalea[a] and Piervincenzo Rizzo
NDE & Structural Health Monitoring Laboratory, Department of Structural Engineering, University of California, San Diego, 9500 Gilman Drive, M.C. 0085, La Jolla, California 92093-0085

Alessandro Marzani[b]
Dipartimento di Strutture, Universita' della Calabria, 87030 Arcavacata di Rende (CS), Italy

(Received 31 March 2003; revised 28 September 2003; accepted 3 October 2003)

This paper deals with the propagation of ultrasonic guided waves in adhesively bonded lap–shear joints. The topic is relevant to bond inspection by ultrasonic testing. Specifically, the propagation of the lowest-order, antisymmetric a_0 mode through the joint is examined. An important aspect is the mode conversion at the boundaries between the single-plate adherents and the multilayer overlap. The a_0 strength of transmission is studied for three different bond states in aluminum joints, namely a fully cured adhesive bond, a poorly cured adhesive bond, and a slip bond. Theoretical predictions indicate that the dispersive behavior of the guided waves in the multilayer overlap is highly dependent on bond state. Experimental tests are conducted in lap–shear joints by a hybrid, broadband laser/air-coupled ultrasonic setup in a through-transmission configuration. The Gabor wavelet transform is employed to extract energy transmission coefficients in the 100 kHz 1.4 MHz range for the three different bond states examined. The cross-sectional mode shapes of the guided waves are shown to have a substantial role in the energy transfer through the joint. [DOI: 10.1121/1.1630999]

PACS numbers: 43.35.Cg, 43.20.Mv, 43.35.Zc [DEC] Pages: 146–156

I. INTRODUCTION

The quality assessment of adhesively bonded components, that are critical to both civil and aerospace structures, remains a long-standing challenge in the field of nondestructive evaluation (NDE). Ultrasonic testing is one powerful method that has been under development for decades for bond inspection applications (Guyott *et al.*, 1986).

Normal-incidence ultrasonic testing has proven useful to calculate the modulus and the thickness of the bond layer in adhesive joints that, in turn, can be related to the bond cohesive strength (Guyott and Cawley, 1988a, 1988b). The method developed in these references exploited ultrasound spectroscopy principles, specifically through-thickness longitudinal resonances of the multilayer joint structure.

The reflection coefficients of bulk waves at oblique incidence are also known to be sensitive to bond state. A rigorous theoretical analysis for this case was presented by Rokhlin and Wang (1991), where reflection coefficients were predicted during the cure process of an epoxy resin layer between solid semispaces.

The waveguide geometry of the layers comprising the adhesive joint lends itself to the propagation of guided waves. The added advantage is the possibility to inspect a long length of the test piece at once. Typically, investigations of guided waves for bond inspection have focused on the multilayer wave propagation problem where the wave is generated and detected in the bonded region. Guided wave amplitude and velocities are among the features sensitive to the elastic properties of the bond layer (Pilarski, 1985; Mal, 1988; Mal *et al.*, 1990; Xu *et al.*, 1990; Pilarski and Rose, 1992; Lowe and Cawley, 1994; Kundu and Maslov, 1997; Chimenti, 1997; Kundu *et al.*, 1998; Rose *et al.*, 1998; Heller *et al.*, 2000; Cheng *et al.*, 2001; Seifried *et al.*, 2002). Another advantage of guided wave inspection is the possibility of maximizing the sensitivity to various types of bond conditions by exciting the appropriate cross-sectional mode shapes through selection of the corresponding frequency–velocity combinations. For example, those modes with large shear stresses at the bond layer, occurring generally for the antisymmetric modes, offer highest sensitivity to disbond-type defects and, generally, to bonded areas with reduced shear stiffness.

When access to the free portion of the individual adherents is allowed, such as in the case of lap–shear joints, bond inspection can be conducted by exciting a Lamb wave in one adherent and receiving the wave in the other adherent across the bondline (long-range method). Essentially, the concept exploits the leakage of the guided wave from one adherent to the other one through the bonded overlap. Various features of the transmitted wave, including phase velocity and amplitude, can be used as indicators of bond quality (Rokhlin, 1991; Rose *et al.*, 1995; Lanza di Scalea *et al.*, 2001). This method has also proven effective for the determination of geometrical dimensions of the bondline (Lowe *et al.*, 2000).

The long-range testing configuration complicates the wave propagation problem due to mode conversions of the Lamb wave entering and leaving the overlap. The complica-

[a]Author to whom correspondence should be addressed. Electronic mail: flanza@ucsd.edu
[b]Research performed as a visiting scholar at the University of California, San Diego.

tion arises from the different dispersive behavior of a single plate when compared to that of the multilayer overlap. The topic was studied analytically by Rokhlin (1991) through a Wiener–Hopf method and numerically by Lowe *et al.* (2000) through finite-element modeling. Rokhlin (1991) predicted the energy transmission coefficients across a rigid bond and across a slip bond. He then successfully monitored the cure of an epoxy bond in an aluminum lap–shear joint by measuring changes in the transmitted wave phase velocity and amplitude. The focus of Rokhlin's experimental work was the lowest-order, symmetric incident Lamb mode s_0 that was examined at various excitation frequencies. The numerical work by Lowe *et al.* (2000) dealt with the determination of overlap length and bond thickness as a function of the wave transmission coefficients across the joint. Transmission coefficients for the incident s_0, a_0, and a_1 modes were predicted as a function of bond geometry. Predictions on mode conversion effects at the overlap edges were also made for the case of the s_0 mode incident on a fully cured bond.

Building on the studies by Rokhlin (1991) and Lowe *et al.* (2000), the present investigation examines the interaction of the lowest-order, antisymmetric Lamb mode a_0 with aluminum lap–shear joints. The focus of the study is to examine the strength of transmission of a_0 as a function of different bond states and in the light of mode conversions in the overlap. The a_0 mode is particularly attractive in ultrasonic NDE applications because of the associated predominance of out-of-plane displacements over in-plane displacements. The opposite is true for the symmetric mode s_0 at low frequencies. It is generally easier to generate and detect out-of-plane displacements with conventional ultrasonic transduction methods.

Three different bond states are examined, namely a ''fully cured'' bond, a ''poorly cured'' bond, and a ''slip'' bond. The dispersive behavior in the overlap is first modeled as a multilayer wave propagation problem for the three bond states by using known matrix-based methods. An experimental investigation based on a hybrid, noncontact ultrasonic system is then presented. The experimental setup uses a pulsed laser for wave generation and air-coupled transducers for wave detection. The laser source generates broadband signals while essentially allowing for single-mode (a_0) excitation in the configuration employed. Transmission coefficient spectra are obtained in a single test for a wide frequency range through a joint time–frequency analysis based on the Gabor wavelet transform. The measured transmission spectra are discussed in terms of mode conversions at the overlap.

II. PROBLEM STATEMENT

The subject problem is schematized in the drawing of Fig. 1(a) showing the essential components of a lap–shear joint. Mode conversion is a consequence of the different dispersive behavior of a single plate (the adherents) when compared to that of a multilayer structure (the bonded overlap). The single-plate modes are herein indicated with a lower case, $s_i(h)$ for the symmetric modes and $a_i(h)$ for the antisymmetric modes, where $i=0,1,2...$ is the mode order and h is the plate thickness. The multilayer modes are indicated

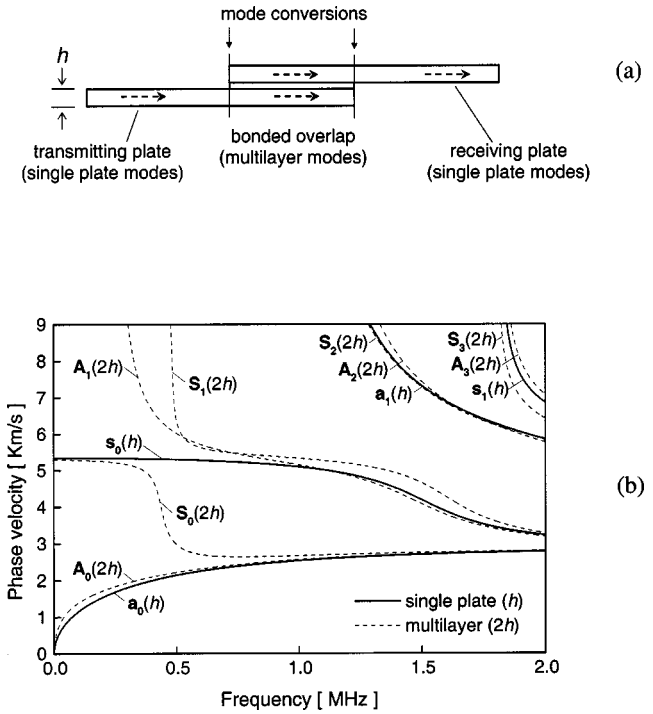


FIG. 1. (a) Schematic of guided wave propagation across a lap–shear joint. (b) Phase velocity dispersion curves for a single 1.524-mm-thick aluminum plate (single plate) and for two, 1.524-mm-thick aluminum plates bonded by a 0.202-mm-thick epoxy layer (multilayer).

with an upper case, $S_i(2h)$ and $A_i(2h)$ for the symmetric modes and the antisymmetric modes, respectively.

Figure 1(b) shows the phase velocity dispersion curves for a single, 1.524-mm-thick aluminum plate superimposed to those of two aluminum plates (1.524-mm-thick each) bonded by a 0.202-mm-thick, fully cured epoxy layer. The ultrasonic properties assumed are shown in Table I. The results for the multilayer case were obtained by a global matrix approach that is discussed in Sec. III. It can be noticed that the multilayer modes shift towards lower frequencies when compared to the single-plate modes. The shift is naturally more pronounced in the highly dispersive branches of the curves. When a Lamb mode is excited in the transmitting plate, it can generally convert to any of the multilayer modes that are supported by the excitation frequencies. The multilayer modes will subsequently convert to single-plate modes when exiting the overlap. For example, Fig. 1(b) shows that for an $a_0(h)$ mode at 0.5 MHz in the transmitting plate, the four multilayer modes $A_0(2h)$, $S_0(2h)$, $A_1(2h)$, and $S_1(2h)$ can potentially be excited in the overlap.

TABLE I. Geometric and ultrasonic properties of the bonded multilayer assumed in the global matrix model.

Layer	Thickness (mm)	Density (kg/m^3)	Bulk long. velocity (km/s)	Bulk shear velocity (km/s)
Aluminum adherent	1.524	2700[a]	6.12[a]	3.10[a]
Fully cured epoxy	0.202	1000[a]	2.20[a]	1.10[a]
Poorly cured epoxy	0.202	1000	1.10	0.55
Water	0.202	1000	1.48	0.000 01

[a]From Lowe *et al.* (2000).

The proportion of conversion into the multilayer modes depends on the wave cross-sectional shapes. Different bond states will produce different mode shapes and, consequently, different transmission strengths for the same incoming mode. It is the dependence of the transmission strength on bond state that is the main focus of this study.

III. MODELLING WAVE PROPAGATION IN THE OVERLAP

A. Basics of the global matrix method

Dispersion curves and cross-sectional mode shapes of the multilayer modes were obtained by the classical global matrix method (GMM) (Knopoff, 1964). An excellent review of matrix-based methods for multilayer wave propagation is given by Lowe (1995).

The overlap region can be modeled as a five-layer structure where the top and bottom layers are half spaces; the remaining three layers represent the two adherent plates and the adhesive layer. It is assumed that the guided waves propagating in each of the layers result from the superimposition of two pairs of partial longitudinal and shear waves, traveling with positive wave numbers and negative wave numbers along the through-the-thickness direction, respectively. A 4×4 field matrix, $[D]$, can be assembled relating the amplitude of the partial waves to the displacement and stress fields in a given layer. By imposing continuity of displacements and stresses at the interfaces between adjacent layers, a 16×20 global matrix can be assembled such that

$$\begin{bmatrix} [D_{1b}] & -[D_{2t}] & & & \\ & [D_{2b}] & -[D_{3t}] & & \\ & & [D_{3b}] & -[D_{4t}] & \\ & & & [D_{4b}] & -[D_{5t}] \end{bmatrix} \begin{bmatrix} [A_1] \\ [A_2] \\ [A_3] \\ [A_4] \\ [A_5] \end{bmatrix} = [0]. \quad (1)$$

The left-hand global matrix in this equation contains the field matrixes relative to the top (subscript t) and the bottom (subscript b) interfaces of each of the five layers. Note that only one of the interfaces is used for the two half-spaces. The column vector contains the amplitudes of the partial waves in the five layers where the two pairs for the ith layer are condensed in the subvector A_i for $i=1,...,5$. Equation (1) represents a homogeneous system of 16 equations in 16 unknown amplitudes once it is assumed that the top and bottom layers are vacuum. By solving the resulting characteristic equation, the dispersive modal solutions for the multilayer are readily obtained.

Calculation of the cross-sectional mode shapes can be carried out once the partial wave amplitudes in all layers are known. The amplitudes are easily calculated from Eq. (1) by assuming known the wave amplitude in one of the layers. In the present study mode shapes were obtained by assuming an incoming longitudinal partial wave with unit amplitude in the top adherent plate.

Damping was not included in the analysis. Damping is generally low for the aluminum adherents, but it can be substantial for the adhesive layer. The expected trend is an increased attenuation of the overlap modes when damping in the adhesive is considered. Since the theoretical portion of this study is limited to dispersion curves and mode shapes, the assumption of loss-free material gives reasonable results.

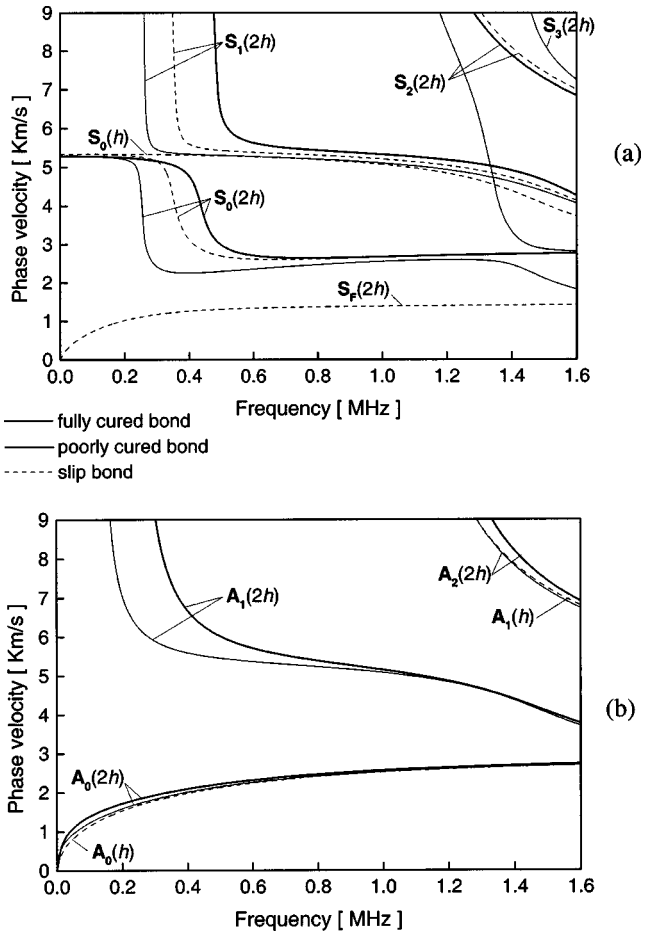


FIG. 2. Phase velocity dispersion curves for two 1.524-mm-thick aluminum plates with the three different interfaces of fully cured epoxy, poorly cured epoxy, and slip (water). Properties in Table I. (a) symmetric modes; (b) antisymmetric modes. Results obtained by the global matrix method.

B. Dispersion results for different bond states

Three different interfaces between two aluminum plates were examined, namely a fully cured epoxy layer, a poorly cured epoxy layer, and a slip (water) interface layer. The geometric and ultrasonic properties assumed for the model are given in Table I. The assumed thickness values were those measured in the specimens tested experimentally. The properties of the aluminum and those of the fully cured epoxy were taken from Lowe *et al.* (2000). The poorly cured epoxy adhesive was simulated by halving the longitudinal and shear bulk velocities of the fully cured epoxy. The longitudinal and shear stiffnesses resulting therefore degraded by a factor of 4, while the density was kept the same as the fully cured epoxy. Finally, nominal values of density and longitudinal velocity were assumed for the water layer. The shear velocity in water was given a small, nonzero value that was required by the numerical analysis.

The multilayer phase velocity dispersion curves obtained by the GMM are shown in Figs. 2(a) and (b) in the

dc–1.6 MHz range for the symmetric modes and for the antisymmetric modes, respectively. Comparing the fully cured bond to the poorly cured bond, in the latter case the highly dispersive branches shift towards lower frequencies and the nondispersive branches shift towards lower phase velocities. A similar trend was found by Cheng *et al.* (2001) comparing a rigid interface to an epoxy (elastic) interface between two aluminum plates. The $S_i(2h)$ modes are known to have large normal stresses and zero shear stresses at the interface layer, resulting in a large sensitivity to the interface longitudinal stiffness. The opposite is true for the $A_i(2h)$ modes and hence the notorious sensitivity to the interface shear stiffness (Kundu and Maslov, 1997). It is therefore not surprising that both the $S_i(2h)$ and the $A_i(2h)$ modes in Figs. 2(a) and (b) are found sensitive to the presence of the poorly cured epoxy where both the longitudinal and the shear stiffness were degraded by the same ratio.

In the case of the slip interface, the $A_i(2h)$ modes cannot propagate due to the inability for the water layer to support shear waves. The symmetric modes $S_i(2h)$ are still supported. Also, the slip multilayer supports the same modes propagating in a single adherent plate (Rokhlin, 1991). These modes are indicated in Fig. 2(a) with $S_0(h)$ and in Fig. 2(b) with $A_0(h)$ for the symmetric and antisymmetric cases, respectively. It can be seen in Fig. 2(a) that the $S_i(2h)$ modes for the slip case are shifted in an intermediate position between the corresponding modes of the two epoxy cases. Given the sensitivity of symmetric modes to the interfacial longitudinal stiffness, the result is expected since the longitudinal bulk velocity assumed for the water layer was between the values taken for the two epoxy cases (Table I).

A mode indicated as $S_F(2h)$ in Fig. 2(a) is found in the slip case at phase velocities below 1.5 km/s. This is the so-called ''fluid mode'' that was first predicted in the numerical analysis by Lloyd and Redwood (1965) for a solid–fluid–solid trilayer structure. The first experimental evidence of the existence of the fluid mode was given more recently by Hassan and Nagy (1997). As shown in Fig. 2(a), $S_F(2h)$ asymptotically tends to the Stoneley velocity at the aluminum–water interface that is slightly smaller than the bulk longitudinal velocity of water. Extraction of displacement mode shapes (see Sec. IV F) confirms the symmetric character of this mode as predicted by Lloyd and Redwood. At low frequencies, the fluid mode is known to result into coupled vibrations of the three constituent layers; at high frequencies, the mode assumes the character of the Stoneley wave with energy concentrated at the solid–fluid interface. The mode shapes calculated for the present configuration confirm that below 300 kHz the $S_F(2h)$ energy is reasonably distributed throughout the thickness of the aluminum adherents. Above 500 kHz, the majority of the energy is found confined to the aluminum–water interfaces with zero energy on the outer edges of the multilayer. Some examples are given in Sec. IV F.

The corresponding group velocity dispersion curves calculated from the phase velocity values are shown in Figs. 3(a) and (b) for the symmetric and the antisymmetric modes, respectively. The group velocity predictions were used to aid in the identification of the overlap modes in the experimental tests.

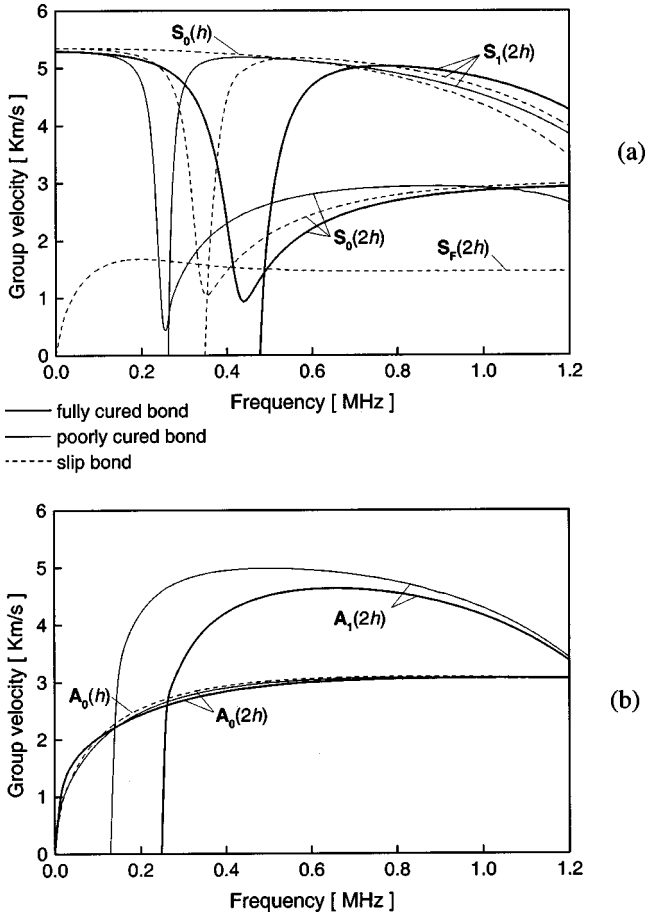


FIG. 3. Group velocity dispersion curves for two 1.524-mm-thick aluminum plates with the three different interfaces of fully cured epoxy, poorly cured epoxy, and slip (water). Properties in Table I. (a) symmetric modes; (b) antisymmetric modes. Results obtained by the global matrix method.

IV. EXPERIMENT: LASER/AIR-COUPLED HYBRID ULTRASONIC TESTS

Through-transmission ultrasonic tests were conducted on aluminum lap–shear joints with three different bond conditions. The goal was to determine the strength of transmission of the $a_0(h)$ Lamb mode generated in one adherent and detected in the other adherent as a function of bond state. The transmission strength was obtained in terms of ''overall'' energy transmission coefficient spectra from the transmitting plate to the receiving plate. ''Intermediate'' transmission coefficient spectra from the transmitting plate to the overlap were also obtained to investigate the role of mode conversions.

A. Specimens

The lap–shear joints examined consisted of 1.524-mm-thick aluminum plates bonded for an overlap length of 62 mm. The principal dimensions are shown in Fig. 4. The adhesive employed was an ambient-cure, Devcon ''2 Ton Epoxy''© type. This was a two-part mixture with a 1:1 resin/

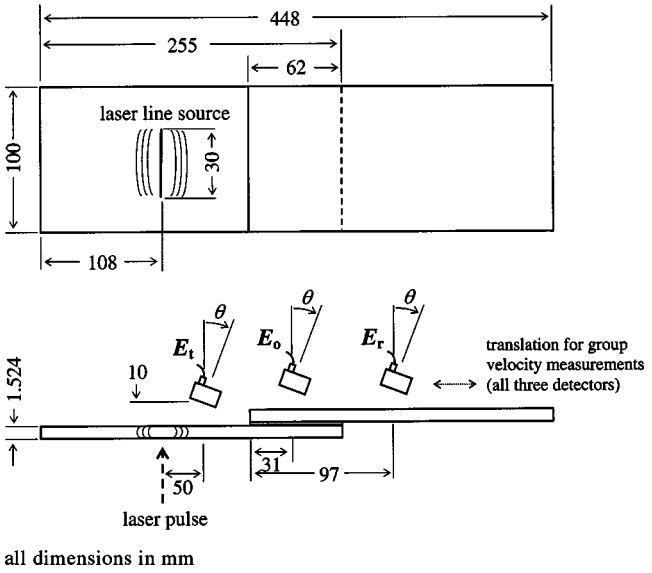


FIG. 4. Principal dimensions of the lap–shear aluminum joints tested experimentally with schematic of the hybrid laser/air-coupled ultrasonic setup for the determination of the transmission coefficient spectra.

hardener volume ratio for full cure with a manufacturer's specified shear strength of 15 MPa and density of 1000 kg/m^3.

A total of three specimens was tested. The first specimen was bonded with the proper resin/hardener ratio of 1:1 resulting in the fully cured interface. The second specimen was bonded with the larger resin/hardener ratio of 1:0.5 resulting in the poorly cured interface. In the third specimen a thin layer of water was interposed in the overlap resulting in the slip interface. The edges of the overlap were sealed with silicone to avoid water leaks. The average bond layer thickness was measured at 0.202 mm (same value assumed in Table I) for the fully cured specimen. The thickness variations for the other two interface layers were measured within 15% of the fully cured value.

B. Experimental setup and procedure

The testing configuration is schematized in Fig. 4. The noncontact arrangement allows for ease of positioning and provides immunity from measurement perturbations due to variations in contact or environmental conditions (coupling medium, contact pressure, temperature). Hutchins *et al.* (1994) have shown that the laser/air-coupled setup is well suited for the study of Lamb mode $a_0(h)$.

The light source was a Q-switched, Nd:YAG pulsed laser operating at 1064 nm with an ~8-ns pulse duration. Through conventional optics the laser beam was focused on the transmitting plate to a 30-mm-long line (Fig. 4). The line source is known to effectively generate directional and broadband signals propagating perpendicular to the line. Ablative generation conditions were produced by applying a Vaseline layer to the irradiated area of the plate that avoided any specimen damage. The strong out-of-plane pulse resulting from the ablative source was most effective for $a_0(h)$ generation.

Micro-machined, air-coupled capacitive transducers were used for signal detection with a 10-mm lift-off distance. These broadband devices had a 10-mm-aperture diameter and a relatively flat frequency response below 1.5 MHz for lift-off distances shorter than 25 mm (Tuzzeo and Lanza di Scalea, 2001).

As traditionally done with conventional wedge transducers, the alignment angle of the air-coupled detectors was adjusted to maximize the sensitivity to a particular guided mode. Owing to Snell's law, $c_p = c_{air}/\sin\theta$, where c_p is the phase velocity of the guided wave, $c_{air} = 0.34$ km/s is the sound speed in air, and θ is the detection angle shown in Fig. 4. As an example, Fig. 5(a) shows an $a_0(h)$ waveform recorded in the transmitting plate at $\theta = 8$ deg. A negligible contribution of $s_0(h)$ can be also seen in the early arrival of the signal. The contribution of $s_0(h)$ remained negligible once the detection angle was decreased to optimize detection of the symmetric mode. The laser line source essentially allowed for single-mode, $a_0(h)$, excitation that was a requirement to obtain transmission coefficients through the joint. In the finite-element analysis by Lowe *et al.* (2000), single-mode excitation was simulated by prescribing a given cross-sectional mode shape for the input signal in the transmitting plate.

The laser-generated signals were detected in the same transmitting plate, in the overlap, and in the receiving plate (Fig. 4). Energy transmission coefficient spectra were obtained as follows:

$$T_{t-r}(f)[\mathrm{dB}] = 10\log\frac{E_r(f)}{E_t(f)}, \tag{2}$$

$$T_{t-o}(f)[\mathrm{dB}] = 10\log\frac{E_o(f)}{E_t(f)}. \tag{3}$$

In Eqs. (2) and (3): T_{t-r} is the overall transmission coefficient from the transmitting plate (t) to the receiving plate (r); T_{t-o} is the intermediate transmission coefficient from the transmitting plate (t) to the overlap (o) as measured from the receiving plate side; E_t, E_r, and E_o are the wave energies detected in the transmitting plate, in the receiving plate, and in the overlap, respectively. Since only two detectors were available, each of the transmission spectra was obtained separately, under a different laser excitation event.

Any geometrical spreading of the laser line-generated ultrasound would produce an underestimation of the transmission coefficients under the present experimental setup. However, spreading would not alter the conclusions on the effect of bond state and propagation frequency on these transmission coefficients. Aindow *et al.* (1982) showed that a 4-mm-long line in aluminum produces a 10° divergence angle (from the normal to the line) of the resulting surface wave. They also found that the divergence angle decreases with increasing length of the line. For the 30-mm-long line used in the present study, beam spreading was thus minimal.

At every measurement point, the detection angle was changed until maximum sensitivity to a particular mode was achieved. Group velocity measurements were also taken for all of the signals to confirm proper identification of the mode on the basis of the theoretical predictions in Fig. 3. Group

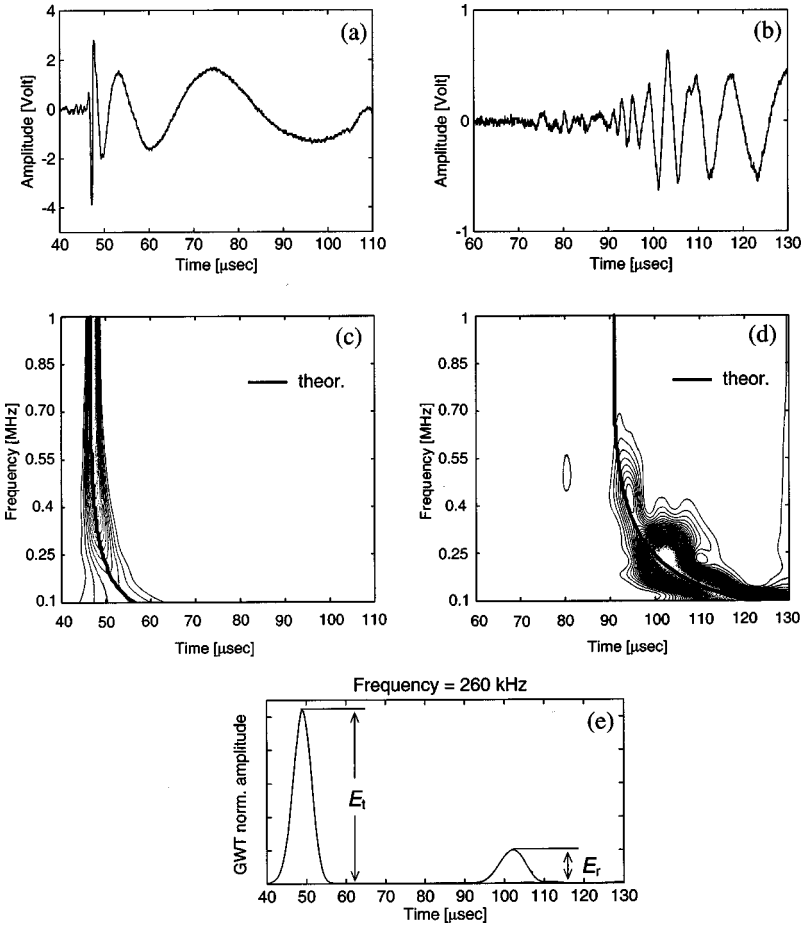


FIG. 5. Laser-generated waveform recorded in the transmitting plate (a) and in the receiving plate across the fully cured bond (b). (c) Gabor wavelet transform (GWT) scalogram of waveform in (a) with theoretical Rayleigh–Lamb solutions for $a_0(h)$. (d) GWT scalogram of waveform in (b) with theoretical Rayleigh–Lamb solutions for $a_0(h)$. (e) Scalogram peaks at 260 kHz for the measurement of the overall transmission coefficient.

velocities were determined by translating each of the detectors by a 15-mm distance with the aid of micrometer-operated stages.

C. Joint time–frequency analysis

A joint time–frequency analysis based on the wavelet transform (WT) was employed for the extraction of energies and group velocities of the dispersive signals. When compared to the 2D Fourier transform method requiring multiple, equally spaced excitation or detection points (Alleyne and Cawley, 1991), the WT only requires a single excitation point and a single detection point. Another feature of the WT is its multiresolution capability deriving from a flexible window that is broader in time for observing low frequencies and shorter in time for observing high frequencies, as required by the Heisenberg uncertainty principle (Kishimoto *et al.*, 1995). Multiresolution is absent in the short-time Fourier transform. When compared to the pseudo-Wigner–Ville distribution, the WT does not generate spurious cross terms that can be detrimental in the presence of multiple echoes (Jeon and Shin, 1993).

The Gabor wavelet transform (GWT) is particularly attractive because it provides the best balance between time and frequency resolution since it uses the smallest possible Heisenberg uncertainty box (Mallat, 1999). The GWT modulates the analyzing wavelet with a Gaussian window $g(t)$ defined as

$$g(t)=\frac{1}{\sqrt[4]{\pi}}\cdot\sqrt{\frac{\eta}{G_S}}\cdot\exp\left(\frac{-(\eta/G_S)^2t^2}{2}\right), \tag{4}$$

where t is time, η is the wavelet center frequency, and G_S is known as the Gabor shaping factor. The choice of η and G_S affects the time–frequency resolution of the analysis. The GWT with $\eta=2\pi$ and $G_S=5.336$ is used in the present study. The value of $\eta=2\pi$ is typically assumed since it conveniently sets the inverse of the wavelet scaling parameter equal to the frequency. The value of $G_S=5.336$ is also common for the study of dispersive wave propagation (see, for example, Kishimoto *et al.*, 1995; Inoue *et al.*, 1996; Gaul *et al.*, 2001; Lanza di Scalea and McNamara, 2003). It should be noted that a different choice of these parameters would alter the absolute magnitude of the GWT scalogram peaks. However, once an acceptable resolution is achieved, the relative magnitude of the peaks taken at the same frequency remains largely unaffected by the particular choice of η and G_S.

The transmission coefficient spectra in Eqs. (2) and (3) were obtained by taking the ratio of the corresponding GWT scalogram peaks at the various frequency values. The procedure is similar to that used to determine ultrasonic frequency-dependent attenuation in a dual detection scheme. Representative results are presented in Fig. 5. This figure shows the application of the GWT to the signals detected in the transmitting plate, Fig. 5(a), and in the receiving plate across the fully cured bond, Fig. 5(b). The two waveforms

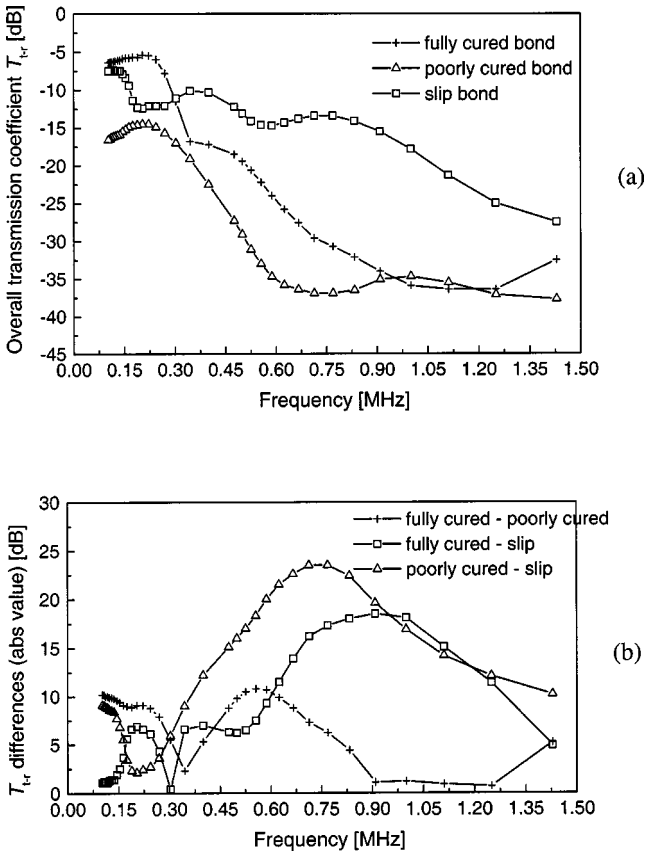


FIG. 6. (a) $a_0(h)-a_0(h)$ overall energy transmission coefficients in aluminum lap–shear joints measured for the fully cured bond, for the poorly cured bond, and for the slip interface. (b) Absolute-valued differences of overall transmission coefficients between different bond states.

were recorded under the same laser excitation event with the two detectors positioned for the measurement of the overall transmission coefficient. The corresponding GWT scalograms are shown in Figs. 5(c) and (d) in the 100 kHz–1 MHz range. The theoretical Rayleigh–Lamb solutions for the $a_0(h)$ mode at the respective detection points are shown superimposed to the scalogram ridges. The comparison confirms that $a_0(h)$ is the only dominant mode both in the transmitting plate and in the receiving plate. This observation was consistent for all of the three bond cases examined. The scalogram peaks for the two signals are plotted in Fig. 5(e) at a frequency of 260 kHz. The peak amplitudes were used to determine the transmission coefficient spectra. Similarly, the peak arrival times at two detection positions were used to extract group velocity values.

D. Overall transmission coefficients

The overall transmission coefficient spectra, $T_{t-r}(f)$, are plotted in Fig. 6(a) as measured for the three bond states examined. The results are $a_0(h)-a_0(h)$ transmission coefficients, that is $a_0(h)$ propagating from the transmitting plate to the receiving plate. It can be seen that the strength of transmission is highly dependent on frequency. The frequency dependence is a function of bond state as well as of attenuation losses of the guided modes. The strongest transmission is measured for the fully cured bond below 300 kHz and for the slip bond above 300 kHz.

Comparisons are made in Fig. 6(b) in terms of absolute-valued differences among the different bond states. These differences indicate the sensitivity of an NDE strategy aimed at discriminating among the different cases based on $a_0(h)$ transmission strength measurements. As nominal values, 9 dB can be taken as a lower boundary for a ''large'' transmission difference and 3 dB as an upper boundary for a ''small'' transmission difference.

Comparing the fully cured bond to the poorly cured bond, the former case presents stronger transmission except in the 900-kHz–1.25-MHz range where the results are comparable. Differences in transmission strength between the two bond states are above 9 dB in the ranges 100–250 kHz and 470–670 kHz. These frequencies are therefore good candidates for an ultrasonic inspection aimed at discriminating between a fully cured and a poorly cured bond. Conversely, differences below 3 dB are found at 345 kHz and in the 900-kHz–1.3-MHz range. These frequencies are therefore not well-suited to discriminate between the two bond states.

Comparing the fully cured bond to the slip bond, differences above 9 dB are found between 580 kHz and 1.3 MHz. Differences below 3 dB are measured below 150 kHz and at 300 kHz. The comparison between the poorly cured bond and the slip bond reveals 345 kHz–1.5 MHz as the most sensitive range (difference larger than 9 dB).

It is also clear in Fig. 5 that marked differences in transmission strength are observed when comparing the slip bond to either one of the adhesive bonds. This is strictly true above 300 kHz, where the slip bond exhibits the strongest transmission of all of the cases examined. As it will be demonstrated in the next section, the strong transmission in the slip case is the result of the overlap mode $A_0(h)$.

The only range where the transmission strength difference is above 9 dB among any two of the three cases examined is limited to 580–670 kHz. This is thus the most sensitive range to discriminate among the three interface conditions at once for the particular specimens examined.

E. Intermediate transmission coefficients

The intermediate transmission coefficient spectra, $T_{t-o}(f)$, are plotted in Fig. 7. The data were obtained by comparing the $a_0(h)$ mode detected in the transmitting plate to the modes detected in the overlap. These results thus indicate the role of the mode-converted overlap signals in the $a_0(h)$ energy transfer mechanism.

Figure 7(a) for the fully cured bond shows that the $a_0(h)$ energy is carried by the overlap modes $A_0(2h)$ and $S_0(2h)$. While $A_0(2h)$ was detected in the entire range 100 kHz–1.4 MHz, $S_0(2h)$ was only activated above 400 kHz. The higher-order overlap modes $A_1(2h)$ and $S_1(2h)$, although theoretically allowed within the frequency range examined (see Fig. 2), were not observed. An explanation for these observations is given on the basis of mode shapes in the following section. It can also be seen in Fig. 7(a) that the majority of the energy transport is due to $A_0(2h)$ below 500 kHz and to $S_0(2h)$ above 500 kHz. $A_0(2h)$ is therefore the overlap mode responsible for the strong overall transmission seen in Fig. 6(a) at low frequencies.

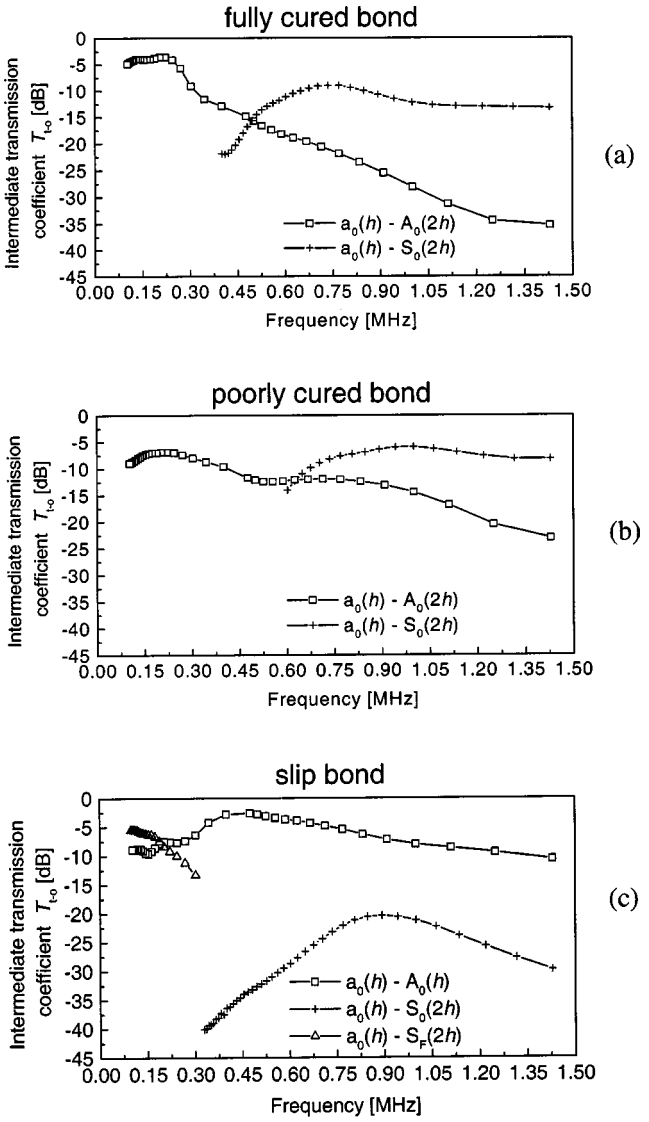


FIG. 7. Intermediate energy transmission coefficients measured in aluminum lap–shear joints for the fully cured bond (a), for the poorly cured bond (b), and for the slip interface (c) showing the role of the overlap modes in the $a_0(h)$ energy transfer mechanism.

Similar results were found in the case of the poorly cured bond shown in Fig. 7(b). The majority of the energy transfer is due to $A_0(2h)$ below 600 kHz and to $S_0(2h)$ above this frequency. The value of 600 kHz is also the cut-on frequency for $S_0(2h)$ that is slightly shifted from the fully cured bond case.

In the slip bond shown in Fig. 7(c), the $A_0(2h)$ mode, not supported by this interface condition, is replaced by $A_0(h)$. The $S_0(2h)$ mode was also measured with a cut-on frequency of 330 kHz. Mode $A_0(h)$ contributes to the majority of the energy transmission. As an example, at 1 MHz $A_0(h)$ carries 16% of the incoming $a_0(h)$ energy. This value is comparable to the 25% $A_0(h)$ contribution predicted by Rokhlin (1991) for a slip overlap. Rokhlin's work, however, calculates the incoming single-plate signal at the overlap edge rather than away from the overlap as done in the present experiment (Fig. 4). The efficiency of $A_0(h)$ in the slip case is, in fact, greater than that of the corresponding $A_0(2h)$ in the epoxy bond cases above 300 kHz. On the contrary, $S_0(2h)$ in the slip case does not propagate as efficiently as in the other two cases. Mode $A_0(h)$ explains the large overall transmission measured in Fig. 6(a) above 300 kHz for the slip bond. Below 300 kHz, the fluid mode, $S_F(2h)$, contributes substantially to the energy transmission. The $S_F(2h)$ transmission strength decreases rapidly with increasing frequency and the mode disappears beyond 300 kHz. This is one of the few instances where experimental proof is given on the existence of the fluid mode.

F. Mode shape considerations

Analysis of the cross-sectional mode shapes can help in understanding some aspects of the mode conversion phenomenon observed in the overlap. As remarked by Auld (1990) and reaffirmed by Lowe *et al.* (2000), an excitation mode will naturally convert to those modes that have similar cross-sectional fields. Consequently, overlap mode shapes most similar to the single-plate $a_0(h)$ shapes are likely to carry most of the energy across the joint. Mode shapes were calculated using the GMM discussed in Sec. III.

Figure 8 presents the in-plane and out-of-plane displacement fields of the fully cured overlap modes along with those of the $a_0(h)$ incoming mode. At 150 kHz, Fig. 8(a), the $A_0(2h)$ shapes match those of $a_0(h)$ while $S_0(2h)$ exhibits a very different symmetry. Consequently, it is solely $A_0(2h)$ that carries the energy through the fully cured bond at these low frequencies as measured in Fig. 7(a). Once the frequency is increased to 500 kHz, Fig. 8(b), the $S_0(2h)$ mode shapes resemble those of the incoming mode; hence, both $A_0(2h)$ and $S_0(2h)$ contribute to the energy transport as confirmed in Fig. 7(a). At 750 kHz, Fig. 8(c), the match between $S_0(2h)$ and $a_0(h)$ is even closer, reflecting the strongest $S_0(2h)$ contribution measured at this frequency in Fig. 7(a). The higher-order modes $A_1(2h)$ and $S_1(2h)$ show symmetries very different from $a_0(h)$. The difference remains throughout the frequency range examined and explains why these modes were never detected in the measurements. This also indicates why, on the contrary, $A_1(2h)$ and $S_1(2h)$ are the primary overlap carriers for the case of incoming $s_0(h)$ mode studied by Lowe *et al.* (2000).

Through similar considerations, the mode shapes obtained for the poorly cured bond, not plotted here, confirm the findings in Fig. 7(b). The main differences with the fully cured case are an increased normal and shear deformation at the bond as a result of the reduced stiffnesses assumed for this layer.

Mode shapes for the slip bond are shown in Fig. 9. At 150 kHz, Fig. 9(a), the $S_F(2h)$ fluid mode and the $A_0(h)$ overlap mode match the incoming $a_0(h)$ as confirmed by the strong transmission measured in Fig. 7(c). The other two modes allowed, $S_0(2h)$ and $S_0(h)$, show different symmetries and hence cannot contribute to the energy transport. At 500 kHz, Fig. 9(b), $S_F(2h)$ concentrates the energy at the aluminum–water interface. The trend becomes more dramatic with increasing frequency as expected from this mode. Thus, $S_F(2h)$ only contributes to the $a_0(h)$ energy transport at low frequencies as indicated in Fig. 7(c). Figure 9(b) also shows why $A_0(h)$ and $S_0(2h)$ were the only modes measured at high frequencies.

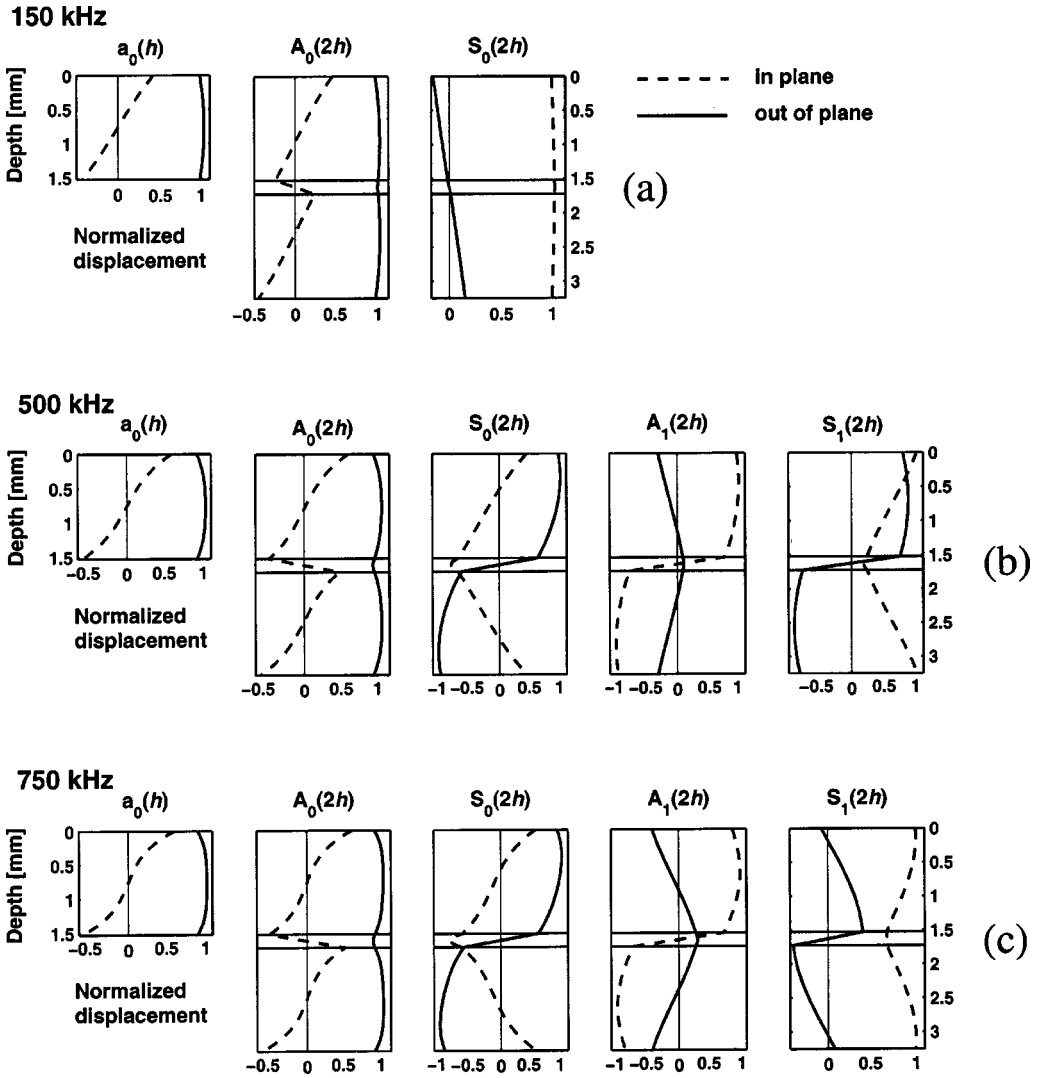


FIG. 8. Cross-sectional displacement mode shapes for two 1.524-mm-thick aluminum plates bonded by a 0.202-mm-thick layer of fully cured epoxy. Propagation frequencies of 150 kHz (a), 500 kHz (b), and 750 kHz (c). Also shown are the corresponding mode shapes for the incoming $a_0(h)$ mode in the single plate.

V. CONCLUSIONS

The broad topic of this paper is the propagation of ultrasonic guided waves in adhesively bonded joints. The specific case examined is the transmission of the lowest-order, antisymmetric Lamb mode $a_0(h)$ across the overlap of lap–shear joints. The subject is relevant to the long-range ultrasonic inspection of bonds. Mode $a_0(h)$ is a good candidate for such an inspection because of the ease of transduction associated with the predominant out-of-plane displacements. Particularly important in this study are mode conversion effects in the overlap region. Three different bond interfaces in aluminum joints are studied, namely a fully cured epoxy adhesive, a poorly cured epoxy adhesive, and a slip interface.

The global matrix method is first employed to model wave propagation in the multilayer overlap. It is shown that phase and group velocity dispersion curves are highly sensitive to bond state. Particularly, the shifts of antisymmetric and symmetric modes are consistent with the relative longitudinal and shear stiffnesses assumed for the three interface layers. The model successfully identifies all of the modes known to propagate in a slip bond including the unique fluid mode.

An experimental investigation involving through-transmission ultrasonic tests by a hybrid laser/air-coupled system is then presented. The setup is particularly well-suited for the study of the $a_0(h)$ mode and yields broadband results. Aluminum lap–shear joints were tested with the three bond states of fully cured epoxy, poorly cured epoxy, and slip (water) interface. The specimens were 1.524-mm-thick plates with an overlap length of 62 mm. The Gabor wavelet transform was employed as the joint time–frequency analysis to extract ultrasonic energy transmission coefficients over the broad frequency range 100 kHz–1.4 MHz.

In terms of ''overall'' $a_0(h)$–$a_0(h)$ transmission coefficients from the transmitting plate to the receiving plate across the overlap, strongest transmission was measured for the fully cured bond below 300 kHz and for the slip interface above 300 kHz. The poorly cured bond resulted in the weakest transmission of all of the cases. Those excitation frequencies that exhibited the largest sensitivity to bond state were

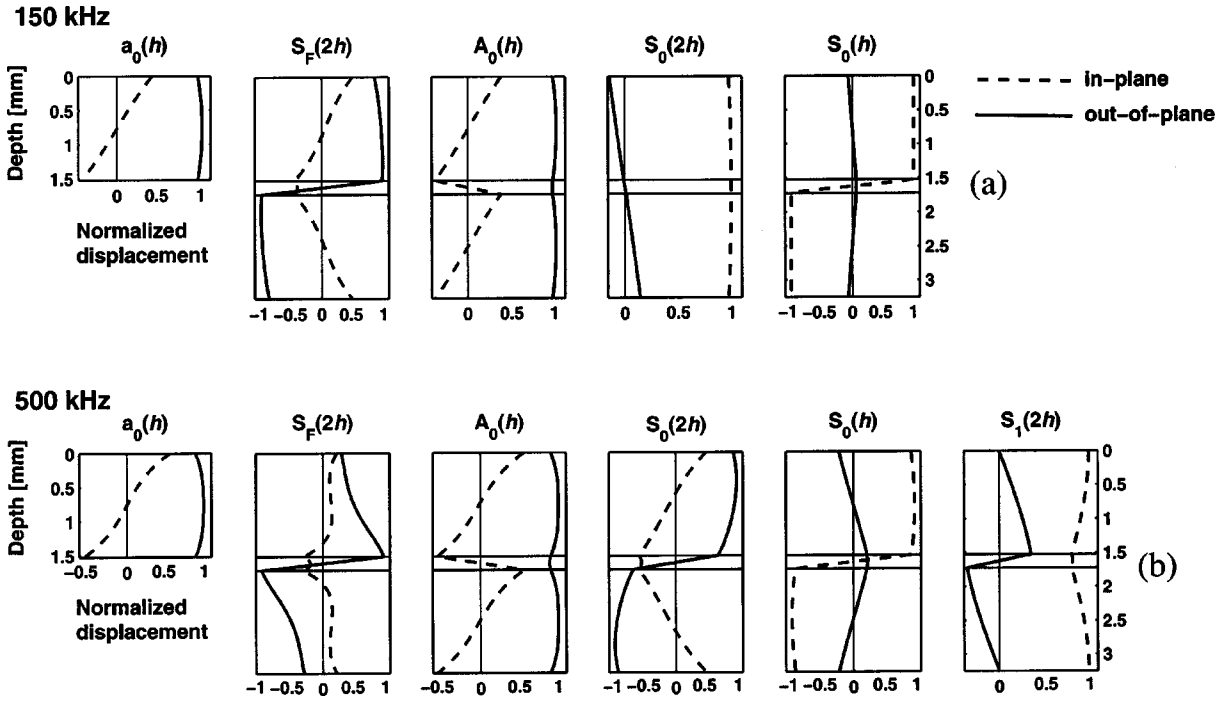


FIG. 9. Cross-sectional displacement mode shapes for two 1.524-mm-thick aluminum plates with an interposed 0.202-mm-thick slip interface. Propagation frequencies of 150 kHz (a) and 500 kHz (b). Also shown are the corresponding mode shapes for the incoming $a_0(h)$ mode in the single plate.

identified as they can be useful for bond inspection. In the frequency range 580–670 kHz, the transmission strength differences among any two of the bond cases were above 9 dB. This range was the most sensitive to bond state for the specimens tested.

Measurement of the ''intermediate'' transmission coefficients from the transmitting plate to the overlap revealed the role of the multilayer modes in the $a_0(h)$ energy transmission. For the fully cured bond, the energy is carried primarily by $A_0(2h)$ below 500 kHz and by $S_0(2h)$ above 500 kHz. A similar behavior was measured for the poorly cured bond, where 600 kHz is the cut-on frequency for $S_0(2h)$. For the slip interface, $A_0(2h)$ is replaced by $A_0(h)$ that contributes to the majority of the energy transmission throughout the 100-kHz–1.4-MHz range. The fluid mode, $S_F(2h)$, also contributes substantially below 300 kHz. The latter finding is particularly interesting since only a few experimental proofs of the existence of this mode exist.

Calculation of the cross-sectional shapes for the overlap modes gives a physical explanation for the experimental findings. The basis for this comparison is that the incoming $a_0(h)$ mode converts to those multilayer modes with similar shapes. The mode shape predictions corroborate this statement by confirming the measured trend of intermediate transmission coefficients.

The results presented strictly apply to the specific specimens tested in this study. A quantitative representation of the variability of the experimental results cannot be given in the lack of data from multiple specimens of each bond type. Besides specimen variability, the inherent shot-to-shot variations of the laser source did not affect the measurements since the transmission coefficients were obtained after normalizing by the incoming energy following Eqs. (2) and (3). Variations in overlap length and bond thickness affect the energy transmission mechanism. The expected trend is an increased transmission with increasing overlap length and bond thickness (Rokhlin, 1991; Lowe *et al.*, 2000). Since cross-sectional mode shapes are what controls the energy transfer mechanism, it would be possible to generalize the role of the overlap modes by accounting for different layer thicknesses.

ACKNOWLEDGMENTS

Partial support for this work was provided by the U.S. National Science Foundation (CMS-0221707) and by the University of California at San Diego Hellman Faculty Fellowship. Alessandro Marzani also acknowledges the Doctoral Fellowship granted by the University of Calabria, Italy.

Aindow, A. M., Dewhurst, R. J., and Palmer, S. B. (**1982**). ''Laser-generation of directional surface acoustic wave pulses in metals,'' Opt. Commun. **42**, 116–120.

Alleyne, D., and Cawley, P. (**1991**). ''A two-dimensional Fourier transform method for the measurement of propagating multimode signals,'' J. Acoust. Soc. Am. **89**, 1159–1168.

Auld, B. A. (**1990**). Acoustic Fields and Waves in Solids, 2nd ed. (Krieger, Malabar, FL).

Cheng, A., Murray, T. W., and Achenbach, J. D. (**2001**). ''Simulation of laser-generated ultrasonic waves in layered plates,'' J. Acoust. Soc. Am. **110**, 848–855.

Chimenti, D. E. (**1997**). ''Guided waves in plates and their use in materials characterization,'' Appl. Mech. Rev. **50**, 247–284.

Gaul, L., Hurlebaus, S., and Jacobs, L. J. (**2001**). ''Localization of a synthetic acoustic emission source on the surface of a fatigue specimen,'' Res. Nondestruct. Eval. **13**, 105–117.

Guyott, C. C. H., and Cawley, P. (**1988a**). ''The ultrasonic vibration characteristics of adhesive joints,'' J. Acoust. Soc. Am. **83**, 632–640.

Guyott, C. C. H., and Cawley, P. (**1988b**). ''Evaluation of the cohesive properties of adhesive joints using ultrasonic spectroscopy,'' NDT Int. **21**, 233–240.

Guyott, C. C. H., Cawley, P., and Adams, R. D. (**1986**). ''The non-destructive testing of adhesively bonded structures: A review,'' J. Adhes. **20**, 129–159.

Hassan, W., and Nagy, P. B. (**1997**). ''On the low-frequency oscillation of a fluid layer between two elastic plates,'' J. Acoust. Soc. Am. **102**, 3343–3348.
Heller, K., Jacobs, L. J., and Qu, J. (**2000**). ''Characterization of adhesive bond properties using Lamb waves,'' NDT & E Int. **33**, 555–563.
Hutchins, D. A., Wright, W. M. D., Hayward, G., and Gachagan, A. (**1994**). ''Air-coupled piezoelectric detection of laser-generated ultrasound,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **41**, 796–804.
Inoue, H., Kishimoto, K., and Shibuya, T. (**1996**). ''Experimental wavelet analysis of flexural waves in beams,'' Exp. Mech. **36**, 212–217.
Jeon, J., and Shin, Y. S. (**1993**). ''Pseudo Wigner–Ville distribution, computer program and its applications to time-frequency domain problems,'' Naval Postgraduate School Report NPS-ME-93-002.
Kishimoto, K., Inoue, H., Hamada, M., and Shibuya, T. (**1995**). ''Time frequency analysis of dispersive waves by means of wavelet transform,'' ASME J. Appl. Mech. **62**, 841–846.
Knopoff, L. (**1964**). ''A matrix method for elastic wave problems,'' Bull. Seismol. Soc. Am. **54**, 431–438.
Kundu, T., Maji, A., Ghosh, T., and Maslov, K. (**1998**). ''Detection of kissing disbonds by Lamb waves,'' Ultrasonics **35**, 573–580.
Kundu, T., and Maslov, K. (**1997**). ''Material interface inspection by Lamb waves,'' Int. J. Solids Struct. **34**, 3885–3901.
Lanza di Scalea, F., Bonomo, M., and Tuzzeo, D. (**2001**). ''Ultrasonic guided wave inspection of bonded lap joints: noncontact method and photoelastic visualization,'' Res. Nondestruct. Eval. **13**, 153–171.
Lanza di Scalea, F., and McNamara, J. (**2003**). ''Measuring high-frequency wave propagation in railroad tracks by joint time-frequency analysis,'' J. Sound Vib. (in press).
Lloyd, P. P., and Redwood, M. (**1965**). ''Wave propagation in a layered plate composed of two solids with perfect contact, slip, or a fluid layer at their interface,'' Acustica **16**, 169–173.
Lowe, M. J. S. (**1995**). ''Matrix techniques for modeling ultrasonic waves in multilayered media,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 525–542.
Lowe, M. J. S., and Cawley, P. (**1994**). ''Applicability of plate wave techniques for the inspection of adhesive and diffusion bonded joints,'' J. Nondestruct. Eval. **13**, 185–200.
Lowe, M. J. S., Challis, R. E., and Chan, C. W. (**2000**). ''The transmission of Lamb waves across adhesively bonded lap joints,'' J. Acoust. Soc. Am. **107**, 1333–1345.
Mal, A. K. (**1988**). ''Guided waves in layered solids with interface zones,'' Int. J. Eng. Sci. **26**, 873–881.
Mal, A. K., Xu, P., and Bar-Cohen, Y. (**1990**). ''Leaky Lamb waves for the ultrasonic nondestructive evaluation of adhesive bonds,'' J. Eng. Mater. Technol. **112**, 255–259.
Mallat, S. (**1999**). *A Wavelet Tour of Signal Processing*, 2nd ed. (Academic, New York).
Pilarski, A. (**1985**). ''Ultrasonic evaluation of adhesive degree in layered joints,'' Mater. Eval. **43**, 765–770.
Pilarski, A., and Rose, J. L. (**1992**). ''Lamb wave mode selection concepts for interfacial weakness analysis,'' J. Nondestruct. Eval. **11**, 237–249.
Rokhlin, S. I. (**1991**). ''Lamb wave interaction with lap-shear adhesive joints: Theory and experiment,'' J. Acoust. Soc. Am. **89**, 2758–2765.
Rokhlin, S. I., and Wang, Y. J. (**1991**). ''Analysis of boundary conditions for elastic wave interaction with an interface between two solids,'' J. Acoust. Soc. Am. **89**, 503–515.
Rose, J. L., Rajana, K. M., and Hansch, M. K. T. (**1995**). ''Ultrasonic guided waves for NDE of adhesively bonded structures,'' J. Adhes. **50**, 71–82.
Rose, J. L., Zhu, W., and Zaidi, M. (**1998**). ''Ultrasonic NDT of titanium diffusion bonding with guided waves,'' Mater. Eval. **56**, 535–539.
Seifried, R., Jacobs, L. J., and Qu, J. (**2002**). ''Propagation of guided waves in adhesive bonded components,'' NDT & E Int. **35**, 317–328.
Tuzzeo, D., and Lanza di Scalea, F. (**2001**). ''Noncontact air-coupled guided wave ultrasonics for detection of thinning defects in aluminum plates,'' Res. Nondestruct. Eval. **13**, 61–78.
Xu, P. C., Mal, A. K., and Bar-Cohen, Y. (**1990**). ''Inversion of leaky Lamb wave data to determine cohesive properties of bonds,'' Int. J. Eng. Sci. **28**, 331–346.

Ultrasonic interferometry for the measurement of shear velocity and attenuation in viscoelastic solids

F. Simonetti and P. Cawley
Department of Mechanical Engineering, Imperial College, London SW7 2AZ, United Kingdom

(Received 24 July 2003; revised 10 October 2003; accepted 13 October 2003)

A method for the measurement of the shear properties of solid viscoelastic materials is presented. The viscoelastic material is cut into a cylindrical sample which is clamped between two rods. The transmission and reflection coefficient spectra of the fundamental torsional mode through the sample are measured by means of two pairs of piezoelectric transducers placed at the free ends of the rod-sample-rod system. Such spectra exhibit maxima and minima which occur approximately at the resonance frequencies of the free viscoelastic cylinder. Therefore, the shear velocity can be obtained by measuring the frequency interval between two consecutive maxima or minima. The shear attenuation is derived by best fitting the analytical expression of the reflection and transmission coefficients to the experimental spectra. The test is very quick to set up as the sample is simply clamped between the two rods. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1631944]

PACS numbers: 43.35.Mr, 43.58.Dj [YHB] Pages: 157–164

I. INTRODUCTION

The measurement of the acoustic shear properties of solid viscoelastic coatings plays a major part in the prediction of the test range of an ultrasonic guided wave propagating in a metallic structure coated with a highly attenuative material such as bitumen.[1,2] Since the properties of viscoelastic materials are extremely sensitive to temperature variations and oxidation phenomena which strongly depend on the surrounding environment, there is a need for a technique which allows rapid *in situ* measurements to be performed. The main difficulty comes from the limited dimensions of samples (typically the size of a coin) due to the breakage of the coating as it is detached from the metallic substrate. Moreover, since for guided wave testing the frequency is well below 100 kHz, only the low frequency properties of the coating are of interest.

A number of ultrasonic techniques based on the response of a sample to an acoustic perturbation have been developed for the assessment of material longitudinal acoustic properties.[3–8] In order to measure the acoustic response, which corresponds to the energy scattered from the sample, either solid or fluid delay lines are usually placed between the sample and the acoustic source/receiver. Measurement of shear properties is more complex. First, shear waves do not propagate in liquid couplants due to the low viscosity; consequently, in order to transmit shear stresses through the contact interface between the excitation source and the sample, rigid coupling is needed. One possibility is to cement the acoustic source onto the sample. However, in order to avoid interference of the acoustic signal with the glue layer, the wavelength of the signal has to be much larger than the glue thickness. Most importantly, chemical reactions between the glue and the sample have to be avoided as they could dramatically change the mechanical properties of the sample. Moreover, such a procedure is time consuming and not suitable for rapid testing. Alternatively, liquid delay lines can be used. In this case, the sample is immersed in a liquid bath and interrogated with an oblique longitudinal wave. The scattered longitudinal field depends on both shear and longitudinal properties of the sample as the incident wave undergoes mode conversion into shear and longitudinal bulk waves within the sample. As a consequence, it is possible to derive the sample properties (shear and longitudinal) from the scattered field through an inversion procedure.[9] However, the accuracy of the measurements decreases when the incident signal wavelength is large compared to the sample dimensions. Solid delay lines are more attractive as they can support shear waves, and the shear properties can directly be derived from the shear scattered field. However, the coupling of the solid buffer with the sample remains an issue.

This paper describes and evaluates a technique for the measurement of the shear properties of viscoelastic solids. A small cylindrical sample is clamped between two buffer rods. Neither glue nor couplant is used at the interfaces between the sample and the two buffers. The fundamental torsional guided mode $T(0,1)$ is sent from the free end of one rod and detected at the free ends of each rod. The shear velocity and shear damping of the sample are derived from the measured reflection and transmission coefficients of $T(0,1)$ through the sample.

II. PREVIOUS WORK

It is known from optics that if a monochromatic plane wave I is incident on a layer that separates two half spaces, it undergoes infinite partial reflections within the layer as shown in Fig. 1. Such reflections cause an interference phenomenon between the front reflection b_0 and the infinite back reflections b_i emerging from the back face of the layer. Similarly, an interference occurs between the wave which penetrates the layer, t_0, and the reflections from the layer front face, t_i. If the incident plane wave is perpendicular to the layer and of unit amplitude, the total reflection R and the total transmission T are[10]

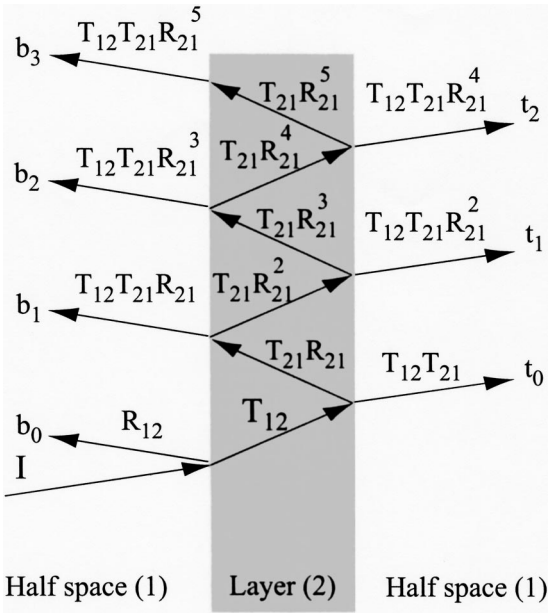


FIG. 1. Schematic diagram of the reflection and transmission of a normally incident wave through a layer separating two half spaces (nonperpendicular incidence is shown for clarity).

$$R = R_{12} + \frac{T_{12}T_{21}R_{21}e^{i2k_2d}}{1 - R_{12}^2 e^{i2k_2d}}, \tag{1}$$

$$T = \frac{T_{12}T_{21}e^{ik_2d}}{1 - R_{12}^2 e^{i2k_2d}}, \tag{2}$$

where the subscripts 1 and 2 refer to the half spaces and the layer, respectively. The coefficients R_{ij} and T_{ij} are the reflection and transmission coefficients at the interface between the half spaces i and j when an incident wave travels from the half space i to the half space j. For acoustic stress waves, by defining the complex impedance Z as[11]

$$Z = \rho \frac{c}{1 + i\alpha c/\omega}, \tag{3}$$

where ρ is the density, ω the angular frequency, c the sound velocity, and α the attenuation in nepers per unit length (c and α are frequency dependent), the reflection and transmission coefficients of either longitudinal or shear waves are

$$R_{ij} = \frac{Z_j - Z_i}{Z_i + Z_j}, \tag{4}$$

$$T_{ij} = \frac{2Z_j}{Z_i + Z_j}. \tag{5}$$

Moreover, k_2 is the wave number within the layer ($k_2 = \omega/c_2 + i\alpha_2$) and d is the layer thickness. By substituting Eqs. (4) and (5) into Eqs. (1) and (2), the total reflection and transmission coefficients can be related to the material impedances:

$$R = \frac{Z_2^2 - Z_1^2}{Z_2^2 + Z_1^2 + 2iZ_2Z_1\cot(k_2d)}, \tag{6}$$

$$T = \frac{4Z_2Z_1}{(Z_1 + Z_2)^2 e^{-ik_2d} - (Z_1 - Z_2)^2 e^{ik_2d}}. \tag{7}$$

In the case of propagation in lossless media, the impedances are real and constant with frequency [see Eq. (3)], and the total reflection and transmission coefficient moduli are periodic functions of the wave number–layer thickness product k_2d. Reflection coefficient modulus maxima [$R = (Z_2^2 - Z_1^2)/(Z_1^2 + Z_2^2)$] occur when

$$\frac{d}{\lambda} = \frac{2N-1}{4}, \quad N \in \{1,2,...\}, \tag{8}$$

where λ is the wavelength within the layer ($\lambda = c_2/f$). On the other hand, the reflection coefficient vanishes when

$$\frac{d}{\lambda} = \frac{N}{2}, \quad N \in \{0,1,2,...\}. \tag{9}$$

Conversely, by energy conservation, transmission coefficient minima and maxima are given by Eqs. (8) and (9), respectively. In other words, at the through thickness resonance frequencies of the free layer [corresponding to condition (9)], the incident wave does not "see" the layer, and all the energy is transmitted without being reflected.

The resonance condition (9) provides two direct methods for the measurement of the velocity c_2. The first method, known as thickness interferometry, is based on the measurement of the reflection or transmission coefficients as a function of the layer thickness d at constant frequency (constant wavelength). The reflection coefficient modulus versus layer thickness, here referred to as the d spectrum, exhibits minima when the thickness is a multiple of half the wavelength according to Eq. (9). Therefore, by measuring the distance between two consecutive minima, Δd, c_2 can be obtained from Eq. (9):

$$c_2 = 2f\Delta d. \tag{10}$$

In a similar fashion, by measuring the transmission d spectrum the velocity can be obtained from the distance between two consecutive peaks using Eq. (10). Thickness interferometry has been employed for the measurement of the longitudinal acoustic properties of gases[12] and several types of liquids such as melts[13–17] and biological tissues *in vitro*.[18,19] The application of this technique has been limited to the case of fluids as the layer thickness can be varied continuously by maintaining perfect coupling between the fluid layer and the two half spaces.

Frequency interferometry is based on the same principle as thickness interferometry. In this case, either R or T is measured as a function of the frequency at constant thickness. The frequencies where reflection minima and transmission maxima occur correspond to the through thickness resonances of the free layer. As a result, the velocity can be obtained by measuring the frequency interval Δf between two consecutive transmission peaks or reflection minima, according to

$$c_2 = 2d\Delta f. \tag{11}$$

When material absorption is considered, due to the fact that the impedances and the wave number k_2 are complex,

the ratios d/λ where maxima and minima of the reflection and transmission coefficients occur are shifted from the values provided by Eqs. (8) and (9). However, the magnitude of the shift, which depends on the amount of internal damping, is usually negligible as discussed by Katahara *et al.*[13] As a result, c_2 can still be derived from Eqs. (10) and (11).

III. METHOD

Frequency interferometry is particularly suitable for the measurement of the acoustic properties of solids as, by contrast with thickness interferometry, there is no need for varying the sample thickness. However, whereas thickness and frequency interferometry are equivalent for a lossless system, a significative difference exists when material damping is considered. In particular, thickness interferometry is a monochromatic method (measurements are performed at constant frequency) and the d spectra do not depend on the frequency dependence of the layer acoustic properties, $\alpha_2(\omega)$ and $c_2(\omega)$. Therefore, at a prescribed frequency ω_0, $R(d)$ and $T(d)$ are known functions of the layer thickness and the two unknown parameters $c_2(\omega_0)$ and $\alpha_2(\omega_0)$ only [see Eqs. (6) and (7)]. Moreover, since $c_2(\omega_0)$ can be obtained from Eq. (10), $\alpha_2(\omega_0)$ can be found by best fitting either $R(d)$ or $T(d)$ to the corresponding experimental d spectra.[13]

In the case of frequency interferometry, such a procedure cannot be employed since the frequency dependence of the acoustic properties $c(\omega)$ and $\alpha(\omega)$ is unknown. As a consequence, the functions $R(\omega)$ and $T(\omega)$ are unknown. In other words, the shape of the reflection and transmission coefficient spectra depends on the shape of the acoustic property spectra. Moreover, for a highly dispersive material, Eq. (11) can no longer be employed as it requires the periodicity of $R(\omega)$ and $T(\omega)$. However, in a sufficiently narrow band of frequency, outside the transition between the material rubbery and glassy behavior,[20] it can always be assumed that the velocity c is constant with frequency and that α is a linear function of the frequency

$$\alpha_2 = \frac{\overline{\alpha_2}}{c_2} f, \tag{12}$$

where the constant $\overline{\alpha_2}$ is the attenuation in nepers per wavelength. Under these assumptions, the problem becomes formally equivalent to thickness interferometry. The material properties, and consequently $R(\omega)$ and $T(\omega)$, are now known functions of the unknown parameters c_2 and $\overline{\alpha_2}$. The velocity c_2 can be obtained from the frequency interval between two consecutive transmission peaks or reflection minima according to Eq. (11) (note that the number of peaks or minima occurring in a given frequency band depends on the layer thickness). The attenuation $\overline{\alpha_2}$ is obtained by best fitting the functions $R(\omega)$ or $T(\omega)$ to the measured spectra.

It has to be emphasized that if the assumptions of constant velocity and constant attenuation per wavelength do not hold in the frequency range of interest the method becomes inaccurate. Nevertheless, the acoustic properties can be found by dividing the frequency range into smaller intervals, where the simple rheological model holds, and increasing the sample length so as to have enough reflection maxima and transmission minima in the interval. In each interval, the validity of this model can be assessed by considering the matching between the experimental spectra and the analytical expressions. As an example, if the shear velocity is highly dispersive, transmission maxima and reflection minima are no longer equally spaced in the frequency domain.

The theory discussed so far refers to the case of plane waves propagating in a boundless space containing an infinite plane layer with finite thickness. Formulas (6) and (7) provide the scattering coefficients for both longitudinal and shear waves when they impinge on the layer perpendicularly. Since the purpose of this paper is to assess the shear properties of the layer, it would be sufficient to measure the reflection and transmission coefficients of a shear bulk wave through the layer rigidly clamped between two large blocks of known properties. However, the propagation of bulk waves is always accompanied by beam spreading due to the finite size of the acoustic source and receiver.

In order to avoid beam spreading guided waves can be employed. The Appendix shows that the reflection and transmission coefficients of the fundamental torsional mode $T(0,1)$ through a disk rigidly clamped between two identical, infinite long rods are still provided by Eqs. (6) and (7). As a result, the shear properties of the disk can be obtained by measuring the reflection and transmission coefficients of $T(0,1)$ following the procedure previously discussed for bulk waves. It is interesting to observe that while $T(0,1)$ and all the other torsional modes do not undergo mode conversion (see the Appendix), all the longitudinal modes propagating above the cutoff frequency of the second longitudinal mode[21] experience mode conversion at the interfaces between the two rods. Moreover, while Eqs. (6) and (7) are exact expressions for the scattering of $T(0,1)$, they represent only approximate solutions for the low frequency scattering of longitudinal modes when the shear velocity c is replaced with the phase velocity of the first longitudinal mode.[22]

The interferometry discussed in this paper reproduces in principle the resonant bar technique.[23,24] In this case, the shear velocity of a rod is measured by fixing two accelerometers at the free ends of the rod and then measuring the frequency interval between two consecutive resonances of the free rod according to Eq. (11). However, the presence of the accelerometers produces a shift of the resonance frequencies which may lead to significant errors especially when the sample dimensions are limited. This problem has been overcome with the split-bar method.[25,26] In this case, the sample is clamped between two rods and its acoustic properties are derived from the resonance frequencies of the rod-sample-rod system. However, the resonances of such a system are different from those of the free sample as they also depend on the rod properties and the accelerometer masses. This is a consequence of the harmonic excitation of the system, which means that the response measured is that of the entire rod-sample-rod system due to the superposition of infinite reflections of the harmonic wave at the free ends of the system. On the other hand, the novelty of this paper is that the two rods work as delay lines which drive an acoustic wave pulse onto the sample. As a consequence, the reflection and transmis-

TABLE I. Material shear properties.

Material	Velocity (m/s)	Attenuation (Np/wavelength)	Density (kg/m^3)
Steel	3260	—	7804
POM	890	0.10	1550
Bitumen	770	0.59	1750

sion through the sample can be completely separated from the reflections at the free ends of the rods and Eq. (11) remains valid. This method is more advantageous than the split-bar method, since the velocity can be derived from the explicit expression (11) rather than inverting more complex resonance expressions as in the case of the split-bar method. Moreover, the method proposed in this paper is more robust because expression (11) does not involve either the properties of the rods or the transduction, which are potential sources of uncertainty.

IV. EXPERIMENTAL SETUP

The setup consisted of two identical stainless steel cylindrical rods of 6 mm diameter and 333 mm long (the acoustic properties are reported in Table I). Alignment of the rods was ensured by two cylindrical supports on each rod, which allowed the rod to slide axially as shown in Fig. 2. Moreover, so as to reduce reflections from the contact interfaces rod-support, the support were made of Nylon 6, whose acoustic impedance is much lower than that of steel. In order to excite and detect the $T(0,1)$ mode, at one end of each rod two piezoelectric transducers were firmly clamped onto the rod lateral surface. The transducers, which were similar to those used for pipe testing,[27] comprised shear elements mounted on a steel backing mass, excited in parallel and oriented as in Fig. 2 to induce torsion. The use of two transducers operating in such a configuration is required in order to avoid the generation of a bending mode which would interfere with $T(0,1)$. A custom-made waveform generator-power amplifier excited the transducers on rod A (see Fig. 2) by a Hanning windowed toneburst. The acoustic response of the system was monitored by receiving the signal with the transducers of rod A operating in pulse-echo mode and those of rod B in pitch-catch mode. The pulse-echo and pitch-catch signals were amplified and transferred to a digital oscilloscope (LeCroy 9400) for digital capture and then stored in a PC.

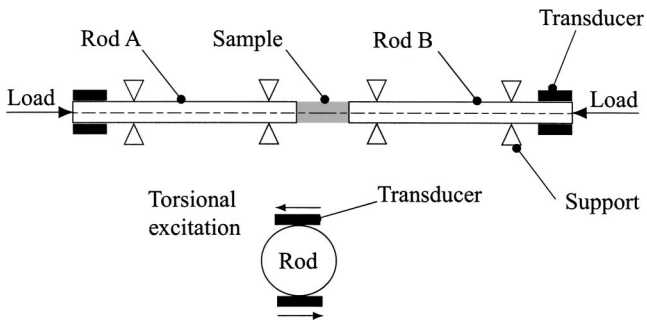


FIG. 2. Diagram of the apparatus for shear property measurements.

According to the theory of linear slip interfaces,[28–30] the dry contact between two surfaces can be modeled by an interfacial spring stiffness, the condition of perfect bonding [Eqs. (A9) and (A10)] being achieved when the spring stiffness is infinite. The stiffness, which depends on the geometry of the contact and on the surface materials, increases under the effect of an applied pressure.[31,32] Therefore, in order to ensure the continuity of displacement and stress at the interfaces between the rods and the disk, an axial compressional load was applied at the free end of each rod by means of a screw.

V. MEASUREMENT PROCEDURE

In order to measure the reflection and transmission coefficients the apparatus was calibrated by performing two types of measurement. At first no sample disk was placed between the two rods which were separated by an air gap. In this configuration the pulse-echo response of rod A (see Fig. 2), $\mathrm{PE}_{\mathrm{air}}$, was recorded. Subsequently, the two rods were put in contact under a compressional load, and the pulse-echo signal of rod A, $\mathrm{PE}_{\mathrm{steel/steel}}$, and the pitch-catch signal $\mathrm{PC}_{\mathrm{steel/steel}}$ were measured. Note that the magnitude of the compressional load is arbitrary since, for the calibration purpose, no intimate contact between the rods is required as discussed later in this section.

Once the calibration measurements were completed, a material sample, previously cut into a 6 mm diameter disk, was placed between the rods. Neither couplant nor glue was employed. The rods were gradually loaded axially and the variations of the transmitted signal continuously monitored on the oscilloscope. The load was increased until the transmitted signal became insensitive to load increments, this condition ensuring intimate contact between the rods and the sample. At that point, the pulse-echo signal of rod A, $\mathrm{PE}_{\mathrm{steel/disk/steel}}$, and the pitch-catch signal $\mathrm{PC}_{\mathrm{steel/disk/steel}}$ were recorded.

As the air impedance is negligible with respect to steel, $\mathrm{PE}_{\mathrm{air}}$ corresponds to unit reflection coefficient; therefore the reflection coefficient R is given by

$$R = \frac{\mathrm{PE}_{\mathrm{steel/disk/steel}}}{\mathrm{PE}_{\mathrm{air}}}. \tag{13}$$

Note that the quantities appearing in the previous relationship and in the rest of this section are the Fourier transforms of the time traces.

The transmission coefficient cannot be calculated by dividing the pitch-catch response $\mathrm{PC}_{\mathrm{steel/disk/steel}}$ by $\mathrm{PE}_{\mathrm{air}}$ as the rod-transducer transfer functions of the two rods do not necessarily coincide due to differences in the transducers and in the coupling. Moreover, the transfer functions of the pulse-echo and pitch-catch electronics are different too. However, it is possible to express the transmission coefficient as

$$T = \eta \frac{\mathrm{PC}_{\mathrm{steel/disk/steel}}}{\mathrm{PE}_{\mathrm{air}}}, \tag{14}$$

where η is a frequency dependent correction factor which can be derived from the pitch-catch response of the rod-rod

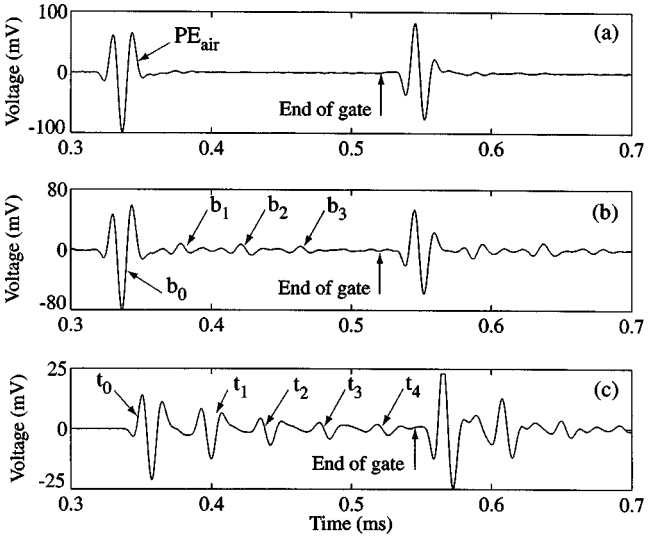


FIG. 3. (a) Typical time-domain signal reflected from the free end of rod *A*; (b) signal reflected at the interface between rod *A* and a 18.5 mm thick POM cylinder clamped between rods *A* and *B*; (c) signal transmitted in the same configuration as (b).

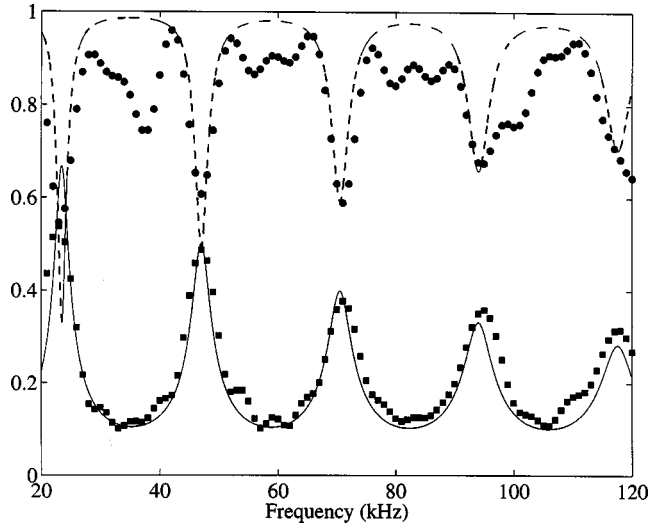


FIG. 4. Measured reflection (●) and transmission (■) coefficient spectra for the POM cylinder clamped between the two rods. (——) transmission coefficient best fit; (— — —) reflection coefficient calculated for $c_2 = 870$ m/s and $\overline{\alpha_2} = 0.105$ Np/wavelength.

contact. In this case, the reflection coefficient at the interface steel-steel, Γ, is

$$\Gamma = \frac{\mathrm{PE}_{\mathrm{steel/steel}}}{\mathrm{PE}_{\mathrm{air}}}. \tag{15}$$

Due to the low frequency considered, it can be assumed that there is no energy dissipation due to internal damping in steel and at the interface between the two rods. As a consequence, by energy conservation, the steel-steel transmission coefficient Λ can be related to Γ according to

$$|\Lambda| = \sqrt{1-|\Gamma|^2}. \tag{16}$$

As a result, since

$$\Lambda = \eta \frac{\mathrm{PC}_{\mathrm{steel/steel}}}{\mathrm{PE}_{\mathrm{air}}}, \tag{17}$$

by equating (16) and (17), it follows that

$$|\eta| = \left| \frac{\mathrm{PE}_{\mathrm{air}}}{\mathrm{PC}_{\mathrm{steel/steel}}} \right| \sqrt{1-|\Gamma|^2}. \tag{18}$$

Thus, η can be found from the calibration measurements with no need to determine the quality of the steel-steel contact in the pitch-catch calibration test. The correction factor so obtained can be substituted into Eq. (14) to provide the transmission coefficient *T*.

VI. RESULTS

The experiments were performed at room temperature 298 K. Figure 3(a) shows the first and second echoes reflected at the free end of rod *A* measured by operating the transducers in pulse-echo mode. Figure 3(b) is the pulse-echo response of rod *A* when a 18.5 mm long acetal copolymer plastic (chemical name polyoxymethylene copolymer, POM), cylinder is clamped between rods *A* and *B*. In this case, each echo consists of a train of signals which contains the front reflection b_0 and the infinite series of back reflections b_i (see Sec. II). The amplitude of the front reflection b_0 is much larger than b_1 since the steel shear impedance is almost 20 times that of POM. Moreover, the amplitude of the back reflections b_i rapidly decays due to energy release into rod *B* and absorption within the POM cylinder. This is more evident from the pitch-catch response shown in Fig. 3(c). In this case also there are two trains of signals. The first train corresponds to the transmitted $T(0,1)$ mode and comprises the signal t_0 and all the reverberations from the front face of the cylinder, t_i (see Sec. II). The second train is the interference of the reverberation of the first train along rod *B* and the signal transmitted by the train reverberating along rod *A*.

The reflection and transmission coefficients were calculated by following the procedure discussed in the previous section. It has to be emphasized that formulas (6) and (7) hold when the steel rods are infinitely long. However, due to the rapid decay of the reverberations within the cylinder, only a finite length of the rods which ensures enough time resolution between the first two trains of signals is required. As a result, for each time trace, the first train of signals was gated out as shown in Fig. 3, and subsequently Fourier transformed. The spectra so obtained were processed according to Sec. V to provide the reflection and transmission coefficient spectra shown in Fig. 4. The transmission coefficient exhibits peaks which occur at equally spaced frequencies $\Delta f = 23.5$ kHz; this confirms that the shear velocity does not change in the frequency range of Fig. 4. From Eq. (11) it follows that the shear velocity is 870 m/s. The same result can be obtained by considering the reflection coefficient minima which occur every 23.5 kHz (Fig. 4). However, it can be observed that the reflection coefficient spectrum contains more noise than the transmission spectrum (Fig. 4). This is due to the fact that, by contrast with the transmitted signal, a large difference between the amplitude of the front reflection b_0 and the back reflections b_i occurs [Figs. 3(b) and (c)]. Such a difference amplifies the effect of noise in the time trace when the Fourier transform is performed. As a

result, in order to derive the attenuation, only the least square fitting (LSF) of Eq. (7) to the measured transmission spectrum was considered. For the POM cylinder, the optimum value of $\overline{\alpha}$ was 0.105 Np/wavelength. The transmission and reflection spectra calculated through Eqs. (7) and (6) for $c_2 = 870$ m/s and $\overline{\alpha}_2 = 0.105$ Np/wavelength are shown in Fig. 4 with solid and dashed lines, respectively. For the transmission coefficient, the fitting is good around the center frequency of the toneburst (60 kHz) and becomes less accurate close to the limits of the pulse bandwidth (around 20 and 110 kHz the pulse energy is low). On the other hand, the noise level of the reflection coefficient spectrum is quite large as mentioned before. The noise can be reduced by using plastic rods so as to reduce the impedance mismatch between the sample and the two rods. In this case, the attenuation can be derived by LSF of Eq. (6) to the experimental reflection coefficient. However, in this paper the analysis is limited to the use of steel rods and the attenuation is derived from the transmission spectrum only.

In order to obtain independent values of the shear velocity and attenuation of POM, a 6 mm diameter 1 m long POM rod was tested. The shear properties were derived from the guided wave attenuation and phase velocity spectra of the $T(0,1)$ mode excited at one end of the POM rod, these spectra corresponding to the dispersion of the shear properties of the rod [see Eq. (A1)]. The $T(0,1)$ mode was monitored by means of laser scanning along the rod axis, $T(0,1)$ being excited with a pair of transducers arranged as discussed in Sec. IV. The signals recorded at different axial positions were Fourier transformed in order to extract the phase velocity and guided wave attenuation spectra as described by Simonetti and Cawley.[33] In the frequency range of interest (between 20 and 100 kHz), both the shear velocity and the attenuation per wavelength ($\overline{\alpha}$) were constant with frequency, the values being summarized in Table I.

The shear velocity obtained from the rod scanning was 2.3% higher than that measured with the frequency interferometry, while the shear attenuation was 5% lower. These differences are around the error levels of the scanning measurements (1% for phase velocity and 10% for attenuation[33]). This suggests that the accuracy of the frequency interferometry method is the same as or even higher than that of the scanning technique. Moreover, the accuracy level of the POM property measurement is affected by the imperfect time resolution between the two trains of signals shown in Fig. 3(c), which produces errors in the evaluation of the transmission coefficient. As a consequence, the previous estimate is conservative when more attenuative materials like bitumen are considered since, in this case, perfect time resolution between the first and second transmitted trains of signals can be achieved.

Bitumen

Since bitumen is softer than POM the condition of good bonding between the disk and the two rods was reached for very low values of the axial compressional load. Figure 5 shows the time traces obtained for a 4.9 mm long bitumen cylinder, the density being 1750 kg/m^3. Note that the reverberations in the steel rods are not visible as the time scale is

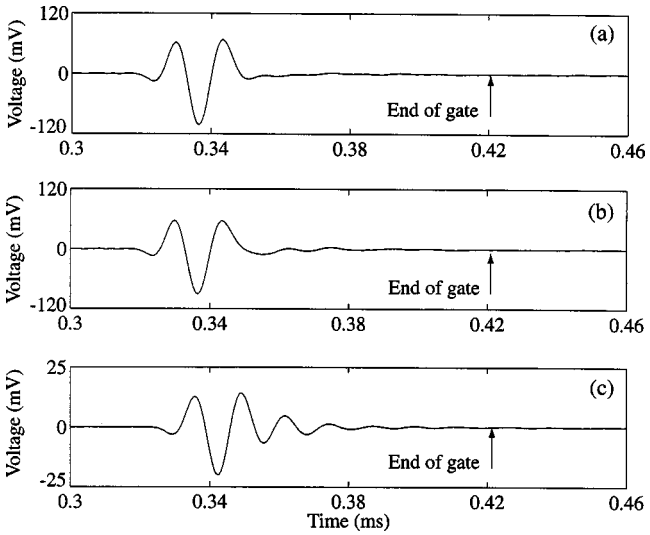


FIG. 5. (a) Time-domain signal reflected from the free end of rod *A*; (b) signal reflected at the interface between rod *A* and a 4.9 mm thick bitumen disk clamped between rods *A* and *B*; (c) signal transmitted in the same configuration as (b).

finer than that of Fig. 3. Figure 5(a) is the pulse-echo response of the free rod *A*. Figure 5(b) shows the reflection at the steel-bitumen interface of rod *A*. Comparison of the time traces (a) and (b) suggests that the reflection coefficient is very little affected by the presence of the bitumen sample since traces (a) and (b) are very similar. Moreover, back reflections cannot be identified due to the low impedance of bitumen and the high attenuation. On the other hand, the transmitted signal [Fig. 5(b)] clearly shows the presence of reverberations from the disk front face t_i. However, these signals are not resolved from the first signal t_0 due to the small sample thickness. On the other hand, a perfect resolution between the first and second trains of signals is achieved. This allows the transmitted signal to be completely separated from the reverberations inside the steel rods, so improving the accuracy of the transmission coefficient measurement.

Figure 6 shows the transmission coefficient spectrum de-

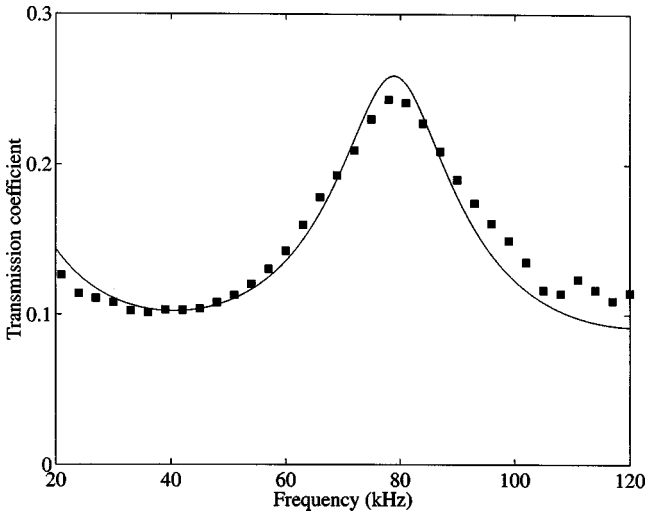


FIG. 6. (■) Measured transmission coefficient spectrum for the bitumen disk; (——) best fit.

rived from the time traces of Fig. 5. The transmission peak shown in Fig. 6, which occurs at 79 kHz, is the second peak as the first peak occurs where the frequency vanishes [see Eq. (7)]. Therefore, $\Delta f = 79$ kHz, and by applying Eq. (11) the corresponding shear velocity is 770 m/s. The shear attenuation, obtained by LSF of Eq. (7) to the measured transmission coefficient, is 0.591 Np/wavelength, which is much larger than that of POM, as expected since bitumen is an extremely attenuative material.[33] The good agreement between the experimental points and Eq. (7) indicates that the shear velocity and shear attenuation $\underline{\alpha}$ are roughly constant when the frequency ranges between 20 and 120 kHz.

VII. CONCLUSIONS

A technique for the measurement of the shear properties of solid viscoelastic materials has been studied and its sensitivity estimated. It has been shown that the shear velocity of a sample clamped between two rods can be obtained by measuring either the reflection or transmission coefficient spectra of the fundamental torsional mode through the sample. The shear attenuation has been derived through a least square fitting of the analytical expression of the transmission coefficient to experimental data.

The good level of accuracy obtained for the shear properties of acetal copolymer plastic shows that a ''rigid'' contact between the sample and the two rods can be achieved by applying a moderate compressional load at the free ends of each rod. As a consequence, since there is no need for the sample to be glued between the rods, the technique is particularly advantageous for rapid measurements. Moreover, as for a lossless sample at the through thickness resonances all the energy passes through the sample without being reflected, irrespective of the impedance difference between the sample and the rods, this technique is particularly suitable for the measurement of very low impedance materials provided the attenuation is low enough.

APPENDIX: SCATTERING OF THE $T(0,1)$ MODE

This appendix studies the scattering of the $T(0,1)$ mode by a disk rigidly clamped between two identical rods.

It can be observed that by performing a transformation from Cartesian to cylindrical coordinates a shear bulk wave transforms into a torsional wave. For an isotropic and homogeneous medium and a cylindrical coordinate system (x,r,ϑ), $\hat{x}$ being coincident with the propagation direction, a plane shear wave becomes the fundamental torsional mode $T(0,1)$, whose nonzero components of displacement and stress fields are[22]

$$u_{\vartheta}(x,r,\vartheta) = rAe^{-\alpha x}e^{i(\omega/c)x}, \tag{A1}$$

$$\tau_{\vartheta x}(x,r,\vartheta) = \mu\left(-\alpha + i\frac{\omega}{c}\right)rAe^{-\alpha x}e^{i(\omega/c)x}, \tag{A2}$$

where A is the wave amplitude and μ the shear modulus. As the stresses τ_{rr}, τ_{rx}, and $\tau_{r\vartheta}$ vanish, it follows that this mode can propagate either in an infinite space or in a free rod whose axis is coincident with the propagation direction $\hat{x}$. Moreover, when two different rods 1 and 2 with the same diameter are rigidly bonded together forming a single infinitely long rod, the $T(0,1)$ mode traveling along rod 1 is scattered at the interface between the two rods into a reflected and a transmitted wave. The displacement fields of the incident, reflected and transmitted waves can be written as

$$u_I = rA_I e^{ik_1 x}, \tag{A3}$$

$$u_R = rA_R e^{-ik_1 x}, \tag{A4}$$

$$u_T = rA_T e^{ik_2 x}, \tag{A5}$$

whereas the stress fields are

$$\tau_I = i\mu_1 k_1 rA_I e^{ik_1 x}, \tag{A6}$$

$$\tau_R = -i\mu_1 k_1 rA_R e^{-ik_1 x}, \tag{A7}$$

$$\tau_T = i\mu_2 k_2 rA_T e^{ik_2 x}. \tag{A8}$$

The reflection and transmission coefficients R_{12} and T_{12} of the stress field are obtained by imposing the continuity of stress and displacement at the interface,

$$u_I + u_R = u_T, \tag{A9}$$

$$\tau_I + \tau_R = \tau_T, \tag{A10}$$

which leads to a system of two equations in the unknown amplitudes A_R and A_T:

$$A_I + A_R = A_T, \tag{A11}$$

$$\mu_1 k_1 (A_I - A_R) = \mu_2 k_1 A_T. \tag{A12}$$

Such a system is identical to that describing the scattering of plane shear horizontal waves at the interface between two different media.[21] Therefore, R_{12} and T_{12} are still provided by Eqs. (4) and (5). It is worth emphasizing that the existence of the solution to the system (A11), (A12) implies that the $T(0,1)$ mode does not undergo mode conversion [i.e., $T(0,1)$ is scattered in the $T(0,1)$ mode only] at the interface between the two rods. This contrasts with $L(0,1)$, where mode conversion to $L(0,2)$ and higher order modes can occur above their cutoff frequencies.

Let us now consider the case of a disk of thickness d rigidly clamped between two identical rods of infinite length. When $T(0,1)$ impinges on the disk, it undergoes infinite reflections as in the case of plane waves. Moreover, since the interface reflection and transmission coefficients R_{ij} and T_{ij} of $T(0,1)$ are the same as those of shear plane waves, the total reflection and transmission coefficients are still given by Eqs. (6) and (7). As a result, the shear properties of a disk clamped between two rods can be obtained by measuring the reflection and transmission coefficients of the fundamental torsional mode following the procedure discussed in Sec. III.

[1] F. Simonetti and P. Cawley, ''On the nature of shear horizontal wave propagation in elastic plates coated with viscoelastic materials,'' submitted to Proc. R. Soc. (2003).

[2] F. Simonetti, ''Lamb wave propagation in elastic plates coated with viscoelastic materials,'' submitted for publication (2003).

[3] D. K. Mak, ''Comparison of various methods for the measurement of reflection coefficient and ultrasonic attenuation,'' Br. J. Non-Destr. Test. **33**, 441–449 (1991).

[4] D. K. Mak, ''Ultrasonic phase velocity measurement incorporating couplant correction,'' Br. J. Non-Destr. Test. **35**, 443–449 (1993).

[5] A. Di Meglio and L. S. Wang, ''A variational method for identification of viscoelastic parameters from experimental data,'' J. Acoust. Soc. Am. **6**, 2476–2753 (2000).
[6] T. Pialucha, C. C. H. Guyott, and P. Cawley, ''Amplitude spectrum method for the measurement of phase velocity,'' Ultrasonics **27**, 270–279 (1989).
[7] N. Guo, M. K. Lim, and T. Pialucha, ''Measurement of the attenuation using a normalized amplitude spectrum,'' J. Nondestruct. Eval. **14**, 9–19 (1995).
[8] W. Sachse and Y. H. Pao, ''On the determination of phase and group velocities of dispersive waves in solids,'' J. Appl. Phys. **49**, 4320–4327 (1978).
[9] M. Castaings, B. Hosten, and T. Kundu, ''Inversion of ultrasonic, plane-wave transmission data in composite plates to infer viscoelastic material properties,'' Nondestr. Test Eval. Int. **33**, 377–392 (2000).
[10] L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1980).
[11] R. M. Christensen, *Theory of Viscoelasticity: An Introduction* (Academic, New York, 1971).
[12] R. C. Williamson and D. Eden, ''Coherent detection technique for variable-path-length measurements of ultrasonic pulses,'' J. Acoust. Soc. Am. **47**, 1278–1281 (1970).
[13] K. W. Katahara, C. S. Rai, M. H. Manghnani, and J. Balogh, ''An interferometric technique for measuring velocity and attenuation in molten rocks,'' J. Geophys. Res. B **86**, 11 779–11 786 (1981).
[14] M. L. Rivers and S. E. Carmichael, ''Ultrasonic studies of silicate melts,'' J. Geophys. Res., B **92**, 9247–9270 (1987).
[15] R. A. Secco, M. H. Manghnani, and T. Liu, ''Velocities and compressibility of komatiitic melts,'' Geophys. Res. Lett. **18**, 1397–1400 (1991).
[16] P. M. Nasch, M. H. Manghnani, and R. A. Secco, ''A modified ultrasonic interferometer for sound velocity measurements in molten metals and alloys,'' Rev. Sci. Instrum. **65**, 682–688 (1993).
[17] N. Yoshimoto, M. Ikeda, M. Yoshizawa, and S. Kimura, ''Sound velocity of molten silicon,'' Physica B **219–220**, 623–625 (1996).
[18] J. Kushibiki, N. Akashi, T. Sannomiya, N. Chubachi, and F. Dunn, ''VHF/UHF range bioultrasonic spectroscopy system and method,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 1028–1039 (1995).
[19] N. Akashi, J. Kushibiki, and F. Dunn, ''Acoustic properties of egg yolk and albumen in the frequency range 20–400 MHz,'' J. Acoust. Soc. Am. **102**, 3774–3778 (1997).
[20] J. D. Ferry, *Viscoelastic Properties of Polymers* (Wiley, New York, 1980).
[21] B. A. Auld, *Acoustic Fields and Waves in Solids*, Vol. 2 (Krieger, Malabar, FL, 1990).
[22] K. F. Graff, *Wave Motion in Elastic Solids* (Clarendon, Oxford, 1975).
[23] J. L. Buchanan, ''Numerical solution for the dynamic moduli of a viscoelastic bar,'' J. Acoust. Soc. Am. **81**, 1775–1786 (1987).
[24] S. L. Garrett, ''Resonant acoustic determination of elastic moduli,'' J. Acoust. Soc. Am. **88**, 210–221 (1990).
[25] R. D. Adams and J. Coppendale, ''Measurement of the elastic moduli of structural adhesives by a resonant bar technique,'' J. Mech. Eng. Sci. **18**, 149–158 (1976).
[26] B. G. Ferguson, ''Calculation of the loss tangent for viscoelastic materials using the triple-bar composite resonance technique,'' J. Acoust. Soc. Am. **76**, 1577–1579 (1984).
[27] 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); **43**, 101 (2001).
[28] H. G. Tattersall, ''The ultrasonic pulse-echo technique as applied to adhesion testing,'' J. Phys. D **6**, 819–832 (1973).
[29] M. Schoenberg, ''Elastic wave behavior across linear slip interfaces,'' J. Acoust. Soc. Am. **68**, 1516–1521 (1980).
[30] M. D. Verweij and H. Chapman, ''Transmission and reflection of transient elastodynamic waves at a linear slip interface,'' J. Acoust. Soc. Am. **101**, 2473–2484 (1996).
[31] B. Drinkwater, R. Dwyer-Joyce, and P. Cawley, ''A study of the transmission of ultrasound across solid-rubber interfaces,'' J. Acoust. Soc. Am. **101**, 970–981 (1996).
[32] A. I. Lavrentyev and S. I. Rokhlin, ''Ultrasonic spectroscopy of imperfect contact interfaces between a layer and two solids,'' J. Acoust. Soc. Am. **103**, 657–664 (1998).
[33] F. Simonetti and P. Cawley, ''A guided wave technique for the characterization of highly attenuative viscoelastic materials,'' J. Acoust. Soc. Am. **114**, 158–165 (2003).

Local surface skimming longitudinal wave velocity and residual stress mapping

Shamachary Sathish,[a),b)] Richard W. Martin,[b)] and Thomas J. Moran
AFRL/Materials and Manufacturing Directorate, AFRL/MLLP, 2230 Tenth Street, Wright-Patterson Air Force Base, Ohio 45433-7817

(Received 22 April 2003; accepted for publication 20 October 2003)

Local variation in surface skimming longitudinal wave (SSLW) velocity has been measured using a scanning acoustic microscope. A very narrow width electrical impulse has been used to excite the transducer of the acoustic lens. This permits the separation of the SSLW signal from the direct reflected signal in the time domain. A simple method of measuring the time delay between the directly reflected signal and the SSLW signal at two defocuses has been utilized for the local measurement of SSLW velocity. The variation in the SSLW velocity measured over an area of the sample is scaled and presented as an image. The method has been implemented to image the variation of the SSLW velocity around a crack tip in a sample of Ti-6Al-4V. Since the SSLW velocity is known to change linearly with the stress, the SSLW velocity image is considered as a representation of the image of stress around the crack tip. Local stress variation in the same region of the crack tip is directly measured using x-ray diffraction. The SSLW velocity image is compared with the x-ray diffraction stress image. The contrast in the two images, spatial resolution, and the penetration depth into the sample of acoustic waves and x rays are discussed. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1631938]

PACS numbers: 43.35.Pt, 43.35.Sx, 43.35.Yb [JGH] Pages: 165–171

I. INTRODUCTION

It has long been known that external loads will induce a small change in the velocity of ultrasonic waves propagating through the material.[1,2] These load-induced sound–velocity changes are characterized mathematically by the incorporation of a parameter known as the acoustoelastic constant in the wave propagation equations. The change in the velocity is linearly proportional to the applied stress and hence has been used effectively in the application of stress measurements using ultrasonic waves.[3,4] The measurement of changes in the velocity with different types of bulk and surface acoustic waves in materials has been utilized for the determination of residual stress. Bulk waves such as longitudinal and shear waves; or surface waves (SAW) such as Rayleigh Waves (RW), Surface Skimming Longitudinal Waves (SSLW), Surface Skimming Shear Waves (SSSW), and Horizontal Shear Waves have been used for that purpose[5–7] (For reviews see Refs. 5–7.)

Most often, the sound velocity measurements are performed by exciting a piezoelectric transducer in contact with a sample and detecting the acoustic waves with another transducer placed at a convenient distance from the transmitter. Since the change in the velocity due to load is on the order of a few tenths of a percent, the acoustic pathlength and the propagation time must be measured with high precision and accuracy. Residual stresses measured using conventional ultrasonic methods are, at best, an average over the propagation distance and over the dimensions of the transducers. These two factors have limited the spatial resolution of the stress measurements to several millimeters.

Acoustic microscope utilizes an acoustic lens, to bring the acoustic beam to focus, to a diffraction-limited spot (a few microns to a few hundred microns of spot size depending on the frequency), in the presence of a coupling fluid like water, on the surface of a sample.[8,9] Scanning acoustic microscopy has provided new methods to measure SAW velocities in a very localized region and thereby allowing the measurement of residual stress with increased spatial resolution. When an acoustic lens is defocused, acoustic rays incident on the water–sample interface at the appropriate critical angles generate both surface skimming longitudinal waves and Rayleigh waves. These SAWs, while propagating at the interface, reradiate energy back into the acoustic lens and can be detected. One of the most used methods of measuring SAW velocity with an acoustic microscope is using the so-called $V(z)$[10,11] curve, which is a record of the amplitude of the signal returning to the transducer after reflection from the water–sample interface, as the distance between the sample and the lens is reduced. Due to the interference between the SAW and the specular reflection at the transducer, the $V(z)$ curve exhibits an oscillatory behavior. The periodicity of the oscillation is directly related to the SAW velocity of the sample. The spatial resolution depends on the frequency of the acoustic wave, the amount of defocus distance, and the type of acoustic lens. A spherical acoustic lens provides an average of the SAW velocity over all propagation directions in the surface of the sample and the resolution depends only on the amount of defocus. A cylindrical acoustic lens can be used to measure the velocity of surface waves propagating in a single direction and is, therefore, very sensitive to material anisotropy.[12] Hence, it has been used very effectively to in-

[a)]Corresponding author. Electronic mail: Shamachary.Sathish@wpafb.af.mil
[b)]Also at the University of Dayton Research Institute, 300 College Park, Dayton, Ohio 45469-0127

vestigate elastic anisotropy. Although the spatial resolution of cylindrical lenses is much lower than that of spherical lenses, only cylindrical lenses can be used to measure the anisotropy of the acoustoelastic constants. SAW velocities have been measured with high precision using a dedicated $V(z)$ curve analysis software package, while collecting the data with a high-precision mechanical displacement system and controlling the temperature of the coupling fluid (usually water) to better than 0.01 °K.[13]

Although the $V(z)$ curve analysis can produce high accuracy and high precision SAW velocity by employing high precision translation stages and controlled coupling fluid temperature, it is very time consuming and computationally intensive. Because of this, it has been used to investigate SAW velocity variations at a few discrete locations[14,15] on specimens. It has never been implemented for mapping residual stress fields or SAW velocity variations on macroscopic scales with microscopic resolution.

In order to overcome some of the limitations and to enhance the accuracy and precision in the local surface wave velocity measurements, Meeks *et al.*[16] developed a high resolution scanning phase measuring acoustic microscope with an acoustic lens capable of producing both longitudinal and shear waves simultaneously. While the longitudinal waves focus on the surface of the sample and provide a constant height reference signal, the mode-converted shear waves produce Rayleigh waves. The phase difference between the two signals is related to the Rayleigh wave velocity. With a lens operating at 670 MHz, it has been possible to map the Rayleigh wave velocity with a resolution of 50 ppm and a spatial resolution of 5–10 μm. Following, Husson's[17] acoustoelastic theory for Rayleigh wave velocity was converted into residual stress. A resolution of 0.33 MPa in residual stress has been achieved. The result of residual stress measurements performed using the high resolution scanning phase measuring acoustic microscope has been found to be in good agreement with the results of a finite element analysis. Such high-resolution stress measurements can be very useful in microelectronics applications. Although the scanning phase measuring microscope has very high resolution, the lens configuration and the electronic instrumentation to measure the phase are quite complex. Recently, Okade and Kawashima[18] have made Rayleigh wave velocity measurements using the $V(z)$-curve method, employing cylindrical lenses with modified instrumentation for defocus-distance determination. These measurements were made to study the stress distributions around holes in aluminum alloys with a spatial resolution of 1 mm. They have measured residual stress at different locations and have compared the measurement results with finite element calculations. Since a cylindrical lens was used and the measurements were performed at some fixed locations, an image of residual stress as produced by Meeks *et al.*[16] has not been generated. Most of the reported stress measurements with acoustic microscopes have been compared with computationally evaluated stress distributions and never with the results of another technique capable of measuring residual stress on the same region of the sample.

In this paper we present a simple method of measuring local surface skimming longitudinal wave velocity, without the need for highly specialized electronics of the scanning phase measuring microscope and dual mode acoustic lens or the time-consuming $V(z)$ curve method. The method has been utilized to measure the variation of SSLW velocity around a crack tip. Based on the linear relationship between the ultrasonic velocity and residual stress, the SSLW velocity variation is interpreted as a scaled image of residual stress around a crack tip. The results of SSLW velocity is compared with an experimentally measured residual stress distribution with x-ray diffraction performed on the same region of the sample. A comparison of the SSLW velocity image and the x-ray stress image is presented. Finally, an analysis of the penetration depth into the sample of the surface skimming longitudinal acoustic waves and the x rays are discussed.

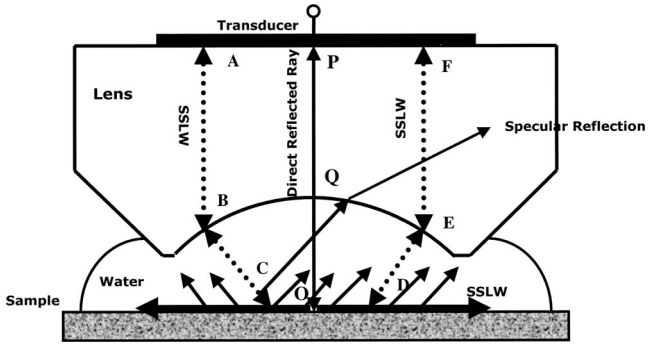


FIG. 1. Important rays in an acoustic lens for the measurement of Surface Skimming Longitudinal Wave (SSLW) velocity.

II. EXPERIMENTAL METHODOLOGY

A. Scanning acoustic microscopy

The principles of measuring the surface wave velocity using the time difference between the directly reflected signal and the SAW signal has been known for quite some time.[19–21] Recently, the sensitivity, accuracy, and the precision of the SAW velocity determination using this method have been analyzed.[22,23] Longitudinal surface skimming waves are generated at the longitudinal critical angle of the solid. This mode-converted wave has been used to measure residual stress in railroad steel structures with large area contact transducers.[24] In acoustic microscopy it has been used to measure acoustoelastic constants using a spherical lens. With cylindrical lenses, anisotropic acoustoelastic constants have been measured on polymeric materials and aluminum.[25,26]

The determination of SSLW velocity with a scanning acoustic microscope is based on the measurement of the time delays of the direct reflected signal (PQ–QO–OQ–QP) and the SSLW signal (AB–BC–CD–DE–EF), as shown in Fig. 1. Following Fig. 1, a relationship between the defocus distance (Z) and the time delay for the SSLW can be derived. The propagation time inside the lens rod for the rays PQ and QP, AB, and DE can be assumed to be the same. The propagation time (t_w) for the direct reflected signal ($t_{QO}+t_{OQ}$) is

$$t_w=\frac{2Z}{V_w}, \tag{1}$$

where V_w is the velocity of sound in water.

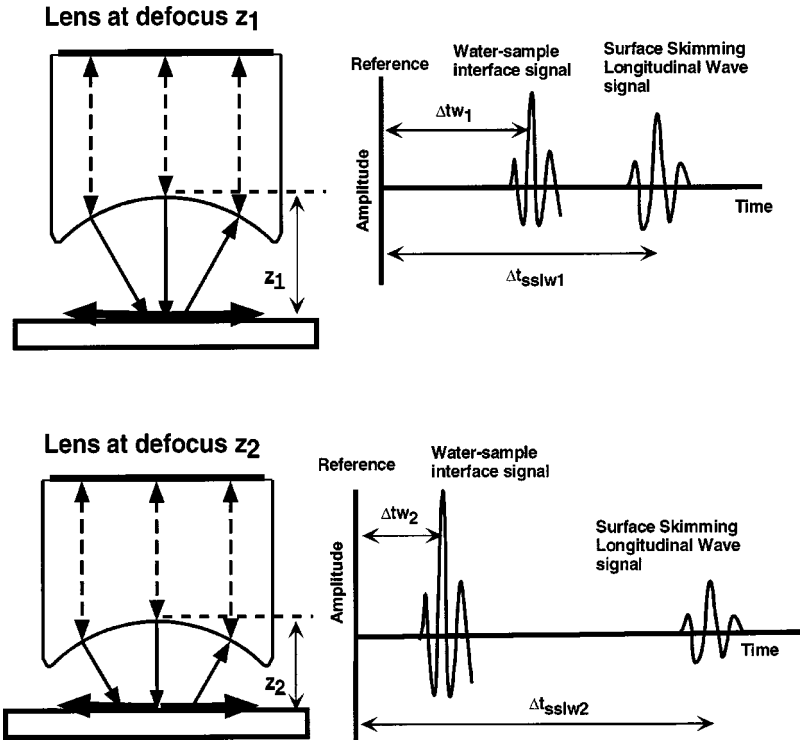


FIG. 2. Time difference measurement procedure for surface skimming longitudinal wave velocity determination with an acoustic lens.

The propagation time (t_{SSLW}) for the SSLW signal ($t_{\mathrm{BC}}+t_{\mathrm{CD}}+t_{\mathrm{DE}}$) is

$$t_{\mathrm{SSLW}}=\frac{2(F-Z/\cos\theta_{\mathrm{SSLW}})}{V_w}+\frac{2Z\tan\theta_{\mathrm{SSLW}}}{V_{\mathrm{SSLW}}}, \tag{2}$$

where V_{SSLW} is the SSLW velocity in the material and $\sin\theta_{\mathrm{SSLW}}=V_W/V_{\mathrm{SSLW}}$.

For the case of two different defocuses Z_1 and Z_2, the difference in propagation time for the direct reflect wave is

$$t_{w1}-t_{w2}=\frac{2(Z_1-Z_2)}{V_w}. \tag{3}$$

Similarly, the difference in propagation time for the SSLW rays is

$$t_{\mathrm{SSLW1}}-t_{\mathrm{SSLW2}}=2(Z_1-Z_2)\left[\frac{1}{V_w\cos\theta_{\mathrm{SSLW}}}-\frac{\tan\theta_{\mathrm{SSLW}}}{V_{\mathrm{SSLW}}}\right]. \tag{4}$$

From these equations a relation between surface skimming longitudinal wave velocity and the time differences can be obtained as

$$V_{\mathrm{SSLW}}=\frac{2(Z_1-Z_2)}{\sqrt{[(t_{w1}-t_{w2})^2-(t_{\mathrm{SSLW1}}-t_{\mathrm{SSLW2}})^2]}}. \tag{5}$$

The principle of the methodology of measuring the SSLW velocity using impulse excitation in an acoustic lens is schematically described in Fig. 2. At a defocus Z_1, the time (Δt_1) between a reference position and the maxima of the direct reflected signal, and the time ($\Delta t_{\mathrm{SSLW1}}$) between the same reference position and minima of the SSLW signal are measured. At another defocus Z_2, the time (Δt_2) between the same reference position and the maxima of the direct reflected signal and the time ($\Delta t_{\mathrm{SSLW2}}$) between the reference and minima of the SSLW signal are measured and stored in the computer. SSLW velocity at the position is determined using Eq. (5), which is modified as

$$V_{\mathrm{SSLW}}=\frac{2(Z_1-Z_2)}{\sqrt{[(\Delta t_1-\Delta t_2)^2-(\Delta t_{\mathrm{SSLW1}}-\Delta t_{\mathrm{SSLW2}})^2]}}. \tag{6}$$

When an acoustic lens is used for local SSLW velocity measurements, the area on the sample in which the velocity is determined depends on both defocus (Z) and the SSLW critical angle. For a spherical acoustic lens, the focusing acoustic beam intersects the flat surface of the sample in a circle. On defocusing the diameter of the circle increases (the radius of the circle$=Z\tan\theta$, where θ is the opening angle of the lens, Fig. 1). During defocusing, at the longitudinal critical angle of the sample, θ_{SSLW}, the SSLW is generated and propagates along the water sample interface. The excitation occurs along a circular path of radius $Z\tan\theta_{\mathrm{SSLW}}$ on the sample and the SSLW propagates toward the center of the circle. This radius is smaller than the diameter of the entire acoustic beam intersecting the sample at that defocus. The reradiated SSLW signal detected by the transducer is an average over the area of this circle on the sample. Hence, even if there are variations in the SSLW velocity in the sample in different directions, a spherical acoustic lens will not be able to detect it, and will only provide an average velocity.

In our measurements a spherical acoustic lens (Panametrics V390) having a focal length of 5 mm, an approximate half-opening angle of 60°, and operating at a nominal frequency of 50 MHz, has been utilized. The transducer is excited using an impulse excitation generated with a 200 MHz pulser/receiver (Panamatrics 5900PR). The received signals were captured using two gigasamples per second sampling Tektronix digitizer (Model #RTD720A) with eight bit resolution. To enhance the signal to noise ratio, the waveforms

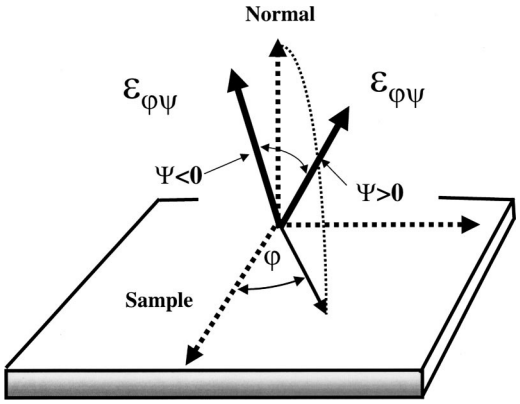


FIG. 3. Sample geometry for x-ray diffraction stress measurements.

are averaged 25 times. A software gating procedure is used for selecting the required portion of the waveform.[27]

At two convenient defocuses the time separation Δt_1, $\Delta t_{\mathrm{SSLW1}}$ and Δt_{w2}, $\Delta t_{\mathrm{SSLW2}}$ are measured using a peak-searching algorithm with special capabilities of reducing noise.[26] The SSLW velocity is computed using Eq. (6). In order to improve the precision the SSLW velocity at the same location was determined several times. The highest and lowest values are discarded, and the final value is obtained as a statistical average. This data was stored in the computer and the acoustic lens is moved to the next position and the measurement procedure is repeated. Finally, the SSLW velocity measurement is scaled and displayed as a gray scale image.[28]

With only two defocus distances, the precision of the SSLW velocity measurement has been determined to be $\pm 0.5\%$ (reference). If time-difference measurements are made at several points during defocus, it might be possible to improve the statistical precision,[29] but the time for acquiring data will also increase.

B. X-ray diffraction stress measurements

Principles of x-ray diffraction for stress measurements have been well established for the last four decades.[30–33] Under a normal unloaded condition, an incident x-ray beam is diffracted by the crystal lattice at angles satisfying the Bragg law ($2d \sin\theta = n\lambda$, where d is the lattice spacing, n $(=1)$, is the order of diffraction, and λ is the wavelength of the x rays). When the material is under load the lattice spacing changes and the angle at which the Bragg diffraction takes place also changes accordingly. A measure of the change in the position of the diffraction peak is the strain caused by the stress in the material. In an isotropic material the strain is related to the stress through the Young's modulus. The stress in the material can be evaluated with the knowledge of Young's modulus and the x-ray diffraction measured strain.

In general, in x-ray diffraction stress measurements for a given orientation of the specimen, the lattice spacing ($d_{\varphi\psi}$) is measured in the direction of the bisector of the incident and diffracted beams, as shown in Fig. 3. Combining Hooke's law with Braag's law, the $\varepsilon_{\varphi\psi}$ can be expressed as

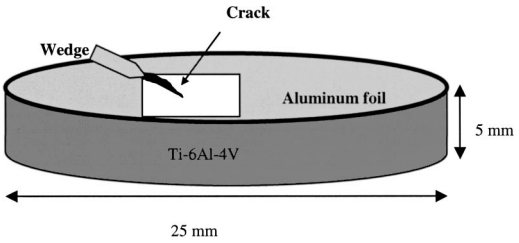


FIG. 4. Cartoon description of the sample, the crack, and the wedge along with the aluminum foil window.

$$\varepsilon_{\varphi\psi} = \left[\frac{d_{\varphi\psi} - d_0}{d_0}\right] = \frac{1}{2} s_2(hkl)[(\sigma_{11}\cos^2\varphi + \sigma_{22}\sin^2\varphi + \sigma_{12}\sin 2\varphi)\sin^2\psi + (\sigma_{13}\cos\varphi + \sigma_{23}\sin\varphi)\sin 2\psi + \sigma_{33}\cos^2\varphi] + s_1(hkl)(\sigma_{11} + \sigma_{22} + \sigma_{33}). \quad (7)$$

In Eq. (7), σ_{ij} are the stress tensor components, $s_1 = -\nu/E$ and $s_2 = (1+\nu)/E$ are the x-ray elastic constants, E is the Young's modulus, and ν is the Possion's ratio of the material. For most practical purposes the penetration depth of x rays is very small, and hence the stress measured can be considered as near surface stress, and hence $\sigma_{33} = 0$ is a fair approximation. For an isotropic material with a biaxial stress condition, the equation (7) simplifies to

$$\left[\frac{d_{\varphi\psi} - d_0}{d_0}\right] = \left[\frac{1+\nu}{E}\right]\sigma \sin^2\psi,$$
$$\sigma = \sigma_{11}\cos^2\varphi + \sigma_{22}\sin^2\varphi, \quad (8)$$

where σ_φ is the biaxial stress.

In an experimental measurement of residual stress using x-ray diffraction, the strain is measured at several ψ angles from a family of crystallographic planes that have a Braag angle in the range of $120 < 2\theta < 180$. The slope of the straight line of the plot of $[(d_{\varphi\psi} - d_0)/d_0]$ vs $\sin^2\psi$ is determined. The residual stress is calculated using the slope of the line, E, the Young's modulus, and ν, the Poisson ration of the material. The diameter of the x-ray spot determines the spatial resolution. To obtain an image of the residual stress, the sample was moved using a computer-controlled motor along the x and y directions and the data were acquired at each location. The collected data was scaled and presented as a gray scale image of residual stress.

C. Sample preparation

The sample used in the study is an alloy of Ti-6Al-4V in the shape of a disk that is 50.4 mm in diameter and 5 mm in thickness (Fig. 4). A radial notch of about 10 mm was cut through the thickness of the sample. A metallic wedge of the same thickness was forced into the notch until a crack appeared at the end of the notch. The sample along with the wedge was metallographically polished on both sides.

In order to perform measurements on the same region with both techniques of scanning acoustic microscopy and x-ray diffraction, a marking procedure for locating the area was developed. A square window ($5\,\mathrm{mm} \times 5\,\mathrm{mm}$) was cut

out of a adhesive aluminum foil; this foil with the cutout was pasted around the crack tip. The edges and the corners of the aluminum foil window were used as reference markers for both acoustic microscopic and x-ray diffraction measurements.

III. RESULTS AND DISCUSSION

A. Surface skimming longitudinal wave velocity

A scanning acoustic microscope with impulse excitation was used to measure the SSLW velocity around the crack tip in the sample over an area of 3 mm$\times$3 mm. For an accurate and precision measurement of SSLW velocity, care should be taken to align the lens and the specimen. In our experimental setup, the lens was mounted on a goniometer capable of tilting in two orthogonal directions. After determining the approximate focus by observing the maximum in the reflected amplitude signal, the lens was tilted back and forth in one direction until a minimum in the time domain is observed. At the same position, the lens is tilted back and forth in the other direction. The adjustment is continued until the reflected signal is at minimum position on the time delay and has maximum amplitude. Finally, the distance between the sample and the lens is varied to observe the maximized reflected signal to identify the focus of the lens. After locating the focus of the lens, the time separation between a reference position and the maximum of the reflected signal is measured. This time corresponds to the total travel time for an acoustic wave generated at the transducer and traveling through the lens, water, reflecting from the surface of the sample, propagating through water and the lens, and returning back to the transducer. For a proper alignment, the sample should be flat and the axis of the lens should be perpendicular to the sample surface. In this condition the total travel time should remain same for all locations in the region of interest on the sample. A map of the total travel time was performed on the region of interest in the sample before attempting SSLW wave velocity measurements. In a uniform region away from the crack over an area of 3 mm $\times$3 mm in the sample, the change in the total travel time was less than 4 ns. This corresponds to a change in pathlength of 3 μm in water. The defocus distance is always measured from the focus of the lens. The time differences Δt_1 and $\Delta t_{\mathrm{SSLW1}}$ were measured at defocus $Z_1=1200$ μm. At another defocus of $Z_2=2200$ μm, the time differences Δt_{w2} and $\Delta t_{\mathrm{SSLW2}}$ were measured. The SSLW is determined using Eq. (5). Moving the sample in steps of 200 μm along the x and y directions, the SSLW velocity measurements were acquired over the entire area. The raw data was mapped to produce an image of the variation of SSLW velocity. Figure 5 shows an image of the variation of SSLW velocity around the crack tip. This image was obtained after smoothing the raw data image with commercial software. In Fig. 5, the crack is marked using a white dashed line. As it can be seen that the SSLW velocity (6200 m/s) is quite high in the region close to the crack tip, while the velocity decreases away from the crack tip to about 5900 m/s. The SSLW wave velocity far from the crack is very similar to the velocity in an unstressed Ti-6Al-4V material.

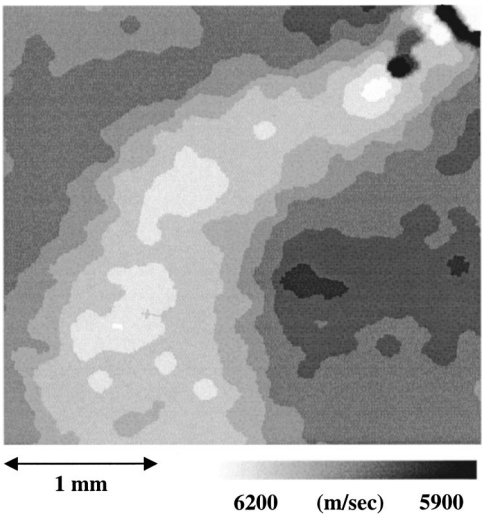


FIG. 5. Surface skimming longitudinal wave velocity variation near the crack tip. The crack tip is located on the top right-hand corner. (The white dashed line indicates the boundary of the crack.)

The SSLW velocity measurements in several materials as a function of external stress have established that the velocity increases linearly with the stress.[6,7] Hence, in comparison with an unstressed condition, in a material under compression the SSLW velocity is higher. Similarly, a material in tension will have a lower SSLW compared to unstressed material. Based on these facts the material in the immediate vicinity of the crack tip, where the SSLW is high (6200 m/s) is expected to be under compression. The material away from the tip in both directions is expected to be tension. These observations are in agreement with well-established facts in fracture mechanics that the material in the vicinity of the crack tip is under compressive stress while the material away from the crack tip is under tensile stress.[34]

The linear relationship between SSLW velocity with stress and the agreement with the fracture mechanics observations encourages us to believe that the SSLW velocity image in Fig. 5 is a scaled image of the stress around a crack tip.

B. X-ray diffraction stress measurements

The x-ray diffraction stress measurements were performed on the same region of the sample where the scanning acoustic microscopy was performed. A titanium x-ray source with the characteristic line K_α ($\lambda=0.2748$ nm) was used for measurement. The aperture and the spot diameter of the x-ray beams were approximately 1 mm. The (302) crystallographic planes of HCP (alpha phase) of Ti-6Al-4V were used as the reference diffraction peak for the stress measurements. The instrument was calibrated using pure titanium powder. Seven sample tilt angles ($\psi=-15°-15°$, steps of 5°) were used for the measurement of changes of lattice spacing; the stress at the location was evaluated by determining the slope of the line given by Eq. (8). In order to average over several grains and improve the signal to noise ratio for a precise peak location, the x-ray source and the detectors were oscillated over 3° at each tilt angle. Scanning the sample in two orthogonal directions in steps of 200 μm, the variation in the

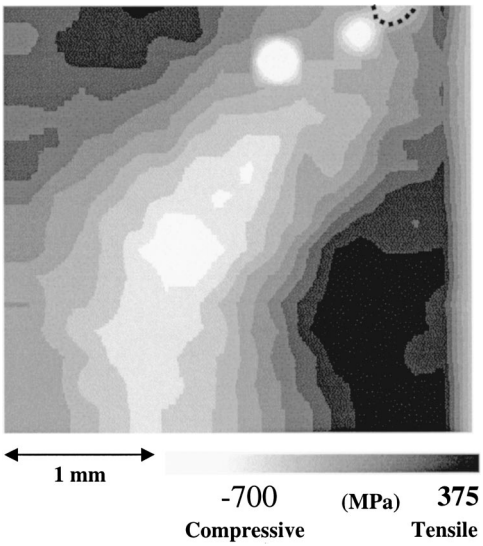


FIG. 6. X-ray diffraction stress variation near the crack tip. (The black dashed line indicates the boundary of the crack.) (The Area of measurement is nearly the same as in Fig. 5.)

stress data was acquired over the same area on which SSLW velocity was measured. The data collected was converted into an image using commercial software. Figure 6 shows a smooth rendered image of the variation of residual stress around the crack tip. It is clearly seen that the stress near the crack tip is highly compressive and is of the order of -700 MPa. On either side away from the crack tip the stress becomes tensile and reaches 375 MPa. As expected, the x-ray stress measurements show that the material in the vicinity of the crack tip is under high compressive stress, reaching the yield strength of the material, while the material away from the crack tip on either side is under tensile stress. This is in agreement with fracture mechanics theoretical calculations[33] and observations with experimental techniques of photoelasticity, Moiré interferometry, Holography, etc. (for reviews of the techniques see Ref. 34).

C. Comparison of SSLW velocity and x-ray diffraction stress images

A comparison of the x-ray stress image with a SSLW velocity image of the same area shows that some features are qualitatively similar, but some differences are also visible. A slight mismatch in the scanned area can be noticed. Along the horizontal direction the x-ray diffraction stress image covers about 300 μm more to the right of the crack, while the SSLW image stops just near the crack. As expected, the SSLW velocity is higher in compressive stress regions than in tensile stress regions. Although the features in the images are qualitatively similar, the width of the central region in x-ray diffraction image is larger than the SSLW velocity image. The SSLW velocity image shows two regions having higher velocity while the x-ray diffraction image shows only one region. One of the possible reasons for the differences is the way stress-measurement data are acquired by the two techniques. In an acoustic microscope, the SSLW velocity measured using a spherical lens is an average over all directions of the circular spot. Hence, any directional dependence of the velocity is averaged out. On the other hand, in the x-ray diffraction the measured stress is along the line of intersection of the tilt circle and the plane of the sample. In our measurements this direction corresponds to a direction perpendicular to the crack. It is well known that the stress along the direction of the crack is different from the stress perpendicular to the crack.[33]

D. Comparison of resolution and penetration depth into the sample in x-ray diffraction and scanning acoustic microscopy

The aperture and the resolution (spot size) of the x-ray diffraction beam are nearly 1 mm. In the case of a scanning acoustic microscope, the two defocuses are used to determine the SSLW velocity. The actual diameters of the acoustic spot at the respective defocuses are 1700 and 3000 μm, based on the half-opening angle of the acoustic lens to be 55°. The corresponding diameters of the circular path along which the SSLW are excited and received at the two defocuses are 600 and 1100 μm. Although the acoustic wave spots are large, the SSLW velocity is evaluated based on the circular path of SSLW excitation.[35,36] Hence, the larger diameter of the circular ring of excitation of the SSLW at the second defocus can be considered as the resolution in these measurements. Thus, the resolution of scanning acoustic microscope SSLW velocity measurements and x-ray diffraction are of the same order.

The x rays penetrate into the surface of the material being examined. Thus, stress measurements performed will be an average over a certain thickness of the material near the surface. The penetration depth of x rays depends on the energy of the source and the other parameters of the material. For the case of a titanium x-ray source impinging a sample of titanium, the penetration depth is about 10–12 μm.[29] Surface acoustic waves are also known to penetrate into the sample. The particle displacement for Rayleigh surface waves decrease exponentially into the depth of the sample and the penetration depth approximately corresponds to a thickness of one Rayleigh wavelength in the material. On the other hand, the determination of the penetration depth of SSLW is rather complicated. Following the theoretical[37] and experimental[38] investigation, the penetration depth for surface skimming waves can be approximately estimated to be about 10–15 μm. This penetration depth is in the same range as that of x rays. Thus, the penetration depth as well as the resolution for both techniques is almost the same and, hence, the images are also similar.

IV. CONCLUSIONS

In conclusion, we have measured the variation of surface skimming longitudinal wave velocity around a crack tip in a sample of Ti-6Al-4V, and compared it with the results of x-ray diffraction stress measurements performed over the same region of the sample. The results from the two measurement techniques are qualitatively similar. The penetration depth of x rays and the penetration depth of surface skimming longitudinal acoustic waves in the material are approximately the same, i.e., about 10–12 μm. The spatial

resolutions of the two techniques are also approximately the same (1 mm). The x-ray diffraction measurements provide an anisotropic stress distribution while the SSLW velocity measurement average out the anisotropy.

ACKNOWLEDGMENTS

The authors acknowledge Tim Campbell for a sample preparation, Mark Ruddell for developing the identification method, Ed Klosterman for help with x-ray measurements, and Brain Frock for a critical reading of the manuscript. The research was sponsored by and performed on-site at the NDE Branch, AFRL/Materials and Manufacturing Directorate, Wright–Patterson Air Force Base, Ohio under Contract No. F33615-98-C-5217.

[1] T. D. Murnaghan, *Finite Deformation of an Elastic Solid* (Wiley, New York, 1951).
[2] D. S. Hughes and J. L. Kelly ''Second-order elastic deformation of solids,'' Phys. Rev. **92**, 1145–1149 (1953).
[3] D. J. Crecraft, ''The measurement of applied and residual stress in metals using ultrasonic waves,'' J. Sound Vib. **5**, 173–193 (1976).
[4] D. M. Egle and D. E. Bray, ''Measurement of acoustoelasticity and third order elastic constants for rail steel,'' J. Acoust. Soc. Am. **60**, 741–744 (1976).
[5] Y. H. Pao, W. Sachse, and H. Fukuoka, ''Acoustoelasticity and ultrasonic measurement of residual stress,'' in *Physical Acoustics*, Vol. XVII, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1984), pp. 62–142.
[6] R. B. Thompson, W.-Y. Liu, and A. V. Clark, Jr., ''Ultrasonic methods,'' in *Handbook of Measurement of Residual Stress*, edited by J. Lu (The Fairmont Press, Lilburn, GA 1996), pp. 149–178.
[7] E. Schneider, ''Ultrasonic techniques,'' in *Structural and Residual Stress Analysis by Nondestructive Methods*, edited by V. Hauk (Elsevier, New York, 1997), pp. 522–563.
[8] C. F. Quate, A. Atlar, and H. K. Wickramasinghe, ''Acoustic microscopy with mechanical scanning: A review,'' Proc. IEEE **67**, 1092–1114 (1979).
[9] A. Briggs, *Acoustic Microscopy* (Clarendon, New York, 1992).
[10] A. Atalar, ''A physical model for acoustic material signatures,'' J. Appl. Phys. **50**, 8237–8239 (1979).
[11] R. D. Weglein, ''A model for predicting acoustic materials signatures,'' Appl. Phys. Lett. **34**, 179–181 (1979).
[12] J. I. Kushibiki and N. Chubachi, ''Material characterization by line-focus beam acoustic microscope,'' IEEE Trans. Sonics Ultrason. **SU-32**, 189–212 (1985).
[13] T. Kobayashi, J. Kushibiki and N. Chubachi, ''Improvement of measurement accuracy of line-focus beam acoustic microscope system,'' 1992 Proceedings of the IEEE Ultrasonics Symposium, 1992, pp. 739–742.
[14] J. I. Kushibiki, H. Takahashi, T. Kobayashi, and N. Chubachi, ''Quantitative evaluation of elastic properties of $LaTiO_3$ crystals by line focus beam acoustic microscopy,'' Appl. Phys. Lett. **58**, 893–895 (1991).
[15] J. I. Kushibiki, H. Takahashi, T. Kobayashi, and N. Chubachi, ''Characterization of $LiNbO_3$ crystals by line focus beam acoustic microscopy,'' Appl. Phys. Lett. **58**, 2622–2624 (1991).
[16] S. R. Meeks, D. Peter, D. Horne, K. Young, and V. Novotny, ''Microscopic imaging of residual stress using a scanning phase-measuring acoustic microscope,'' Appl. Phys. Lett. **55**, 1835–1837 (1989).
[17] D. Husson, ''A perturbation theory for the acoustoelastic effect of surface waves,'' J. Appl. Phys. **57**, 1562–1568 (1985).
[18] M. Okade and K. Kawashima, ''Local stress measurement on polycrystalline aluminum by an acoustic microscope,'' Ultrasonics **36**, 933–939 (1998).
[19] K. Yamanaka, ''Surface acoustic wave measurements using an impulsive converging beam,'' J. Appl. Phys. **54**, 4323–4329 (1983).
[20] R. S. Gilmore, K. C. Tamm, J. D. Young, and D. R. Young, ''Acoustic microscopy from 10 to 100 MHz for industrial applications,'' Philos. Trans. R. Soc. London, Ser. A **320**, 215–235 (1986).
[21] A. Kulik, G. Gremaud, and S. Sathish, ''Direct measurements of the SAW velocity and attenuation using continuous wave reflection scanning acoustic microscope (SAMCRUW),'' in *Acoustical Imaging*, edited by H. Lee and G. Wade (Plenum, New York, 1991), Vol. 18, pp. 227–235.
[22] S. Sathish, R. W. Martin, and T. E. Matikas, ''Rayleigh wave velocity mapping using scanning acoustic microscope,'' in *Review of Progress in Quantitative Nondestructive Evaluation* edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1997), Vol. 18, pp. 2025–2030.
[23] S. Sathish and R. W. Martin, ''Quantitative imaging of Rayleigh wave velocity with a scanning acoustic microscope,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 550–557 (2002).
[24] T. Leon-Salamanca and D. E. Bray, ''Residual stress measurements in steel plates and welds using critically refracted longitudinal (LCR) waves,'' Res. Nondestruct. Eval. **7**, 169–184 (1995).
[25] M. Obata, H. Shimada, and T. Mihara, ''Stress dependence of leaky surface wave on PMMA by line focus beam microscope,'' Exp. Mech. **30**, 34–39 (1990).
[26] Y. C. Lee, J. O. Kim, and J. D. Achenbach, ''Measurement of stresses by line focus acoustic microscopy,'' Ultrasonics **32**, 359 (1994).
[27] R. W. Martin, S. Sathish, and T. E. Matikas, ''Development of a scan system for Rayleigh wave velocity mapping,'' in *Review of Progress in Quantitative Nondestructive Evaluation*, pp. 2031–2038.
[28] S. Sathish and R. W. Martin, ''Development of a scan system for Rayleigh, shear and longitudinal wave velocity mapping,'' *1999 IEEE Ultrasonics Symposium Proceedings*, 1999, pp. 597–600.
[29] D. Xiang, N. N. Hsu, and G. V. Blessing, ''Statistical error analysis of time and polarization resolved ultrasonic measurements,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 1006–1016 (1998).
[30] B. D. Cullity, *Elements of X-Ray Diffraction* (Addison-Wesley, Menlo Park, CA, 1978).
[31] I. C. Noyan and J. B. Cohen, *Residual Stress Measurement by Diffraction and Interpretation* (Springer-Verlag, New York, 1987).
[32] V. Hauk, *Structural and Residual Stress Analysis by Nondestructive Methods* (Elsevier, New York, 1997).
[33] M. Francois, F. Convert, J. M. Sprauel, J. L. Lebrun, C. F. Dehan, R. W. Hendricks, and M. R. James., ''X-ray diffraction method,'' in *Handbook of Measurement of Residual Stress*, edited by J. Lu (The Fairmont Press, Lilburn, GA 1996), pp. 149–178.
[34] C. W. Smith and A. S. Kobayashi, ''Experimental fracture mechanics,'' in *Handbook on Experimental Mechanics*, edited by A. S. Kobayashi (VCH, New York, 1993), pp. 905–968.
[35] I. R. Smith, H. K. Wickramasinghe, G. W. Farnell, and C. K. Jen, ''Confocal surface acoustic wave microscopy,'' Appl. Phys. Lett. **42**, 411–413 (1983).
[36] K. H. Chan and H. L. Bertoni, ''Ray representation of longitudinal lateral waves in acoustic microscopy,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **38**, 27–34 (1991).
[37] L. V. Basatskaya and L. N. Ermolov, ''Theoretical study of ultrasonic longitudinal subsurface waves in solid media,'' Defektoskopiya **7**, 58–65 (1980).
[38] L. A. Nikiforov and A. V. Kharitnov, ''Parameters of longitudinal subsurface waves excited by angle beam transducers,'' Defektoskopiya **6**, 80–85 (1981).

Interaction of a scanning laser-generated ultrasonic line source with a surface-breaking flaw

Younghoon Sohn[a] and Sridhar Krishnaswamy[b]
Center for Quality Engineering and Failure Prevention, Northwestern University Evanston, Illinois 60208

(Received 11 July 2003; revised 3 October 2003; accepted 13 October 2003)

The scanning laser source (SLS) technique has been proposed recently as an effective way to investigate small surface-breaking cracks. By monitoring the amplitude and frequency changes of the ultrasound generated as the SLS scans over a defect, the SLS technique has provided enhanced signal-to-noise performance compared to the traditional pitch–catch or pulse–echo ultrasonic methods. In previous work, either a point source or a short line source was used for generation of ultrasound. The resulting Rayleigh wave was typically bipolar in nature. In this paper, a scanning laser *line* source (SLLS) technique using a true thermoelastic line source (which leads to generation of monopolar surface waves) is demonstrated experimentally and through numerical simulation. Experiments are performed using a line-focused Nd:YAG laser and interferometric detection. For the numerical simulation, a hybrid model combining a mass-spring lattice method (MSLM) and a finite difference method (FDM) is used. As the SLLS is scanned over a surface-breaking flaw, it is shown both experimentally and numerically that the monopolar Rayleigh wave becomes bipolar, dramatically indicating the presence of the flaw. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1630997]

PACS numbers: 43.35.Zc, 43.35.Yb [SFW] Pages: 172–181

I. INTRODUCTION

Surface acoustic waves can be used to detect surface-breaking flaws and are therefore of interest in nondestructive evaluation (NDE). The standard inspection techniques include pulse–echo or pitch–catch methods using contact (typically piezoelectric) transducers for both ultrasound generation and signal detection.[1] These conventional methods have several limitations in practical applications. Contact transducers require the use of a couplant between the transducer and the specimen surface. It is also not possible to use contact methods if the test specimen is at a high temperature or is not stationary (for instance, in process monitoring applications in a production environment). In addition, owing to their size, contact transducers are difficult to use on small objects or on structures with complex geometry.

Laser ultrasonics has the potential to overcome some of these limitations of conventional piezoelectric transducers.[2] Optical generation of acoustic waves using pulsed laser source provides a broadband signal, thereby improving resolution for detection of small cracks. Optical interferometers can also be used as broadband receivers for detection of surface ultrasonic displacements. Since optical generation and detection is noncontact, laser ultrasonics readily lends itself to scanning over large areas, thereby decreasing inspection time. Laser ultrasonics can be used in conventional pulse–echo and pitch–catch modes to detect surface-breaking flaws.[3–6] However, in common with all ultrasonic techniques, when the depth of the crack is much smaller than the wavelength of the incident Rayleigh wave a significant portion of the Rayleigh wave energy passes right under the crack. This leads to quite small changes in observed signal amplitude (for both pulse–echo and pitch–catch methods), making it hard to detect the presence of such small defects.

Recently, a powerful method of detecting surface-breaking cracks, called the scanning laser source (SLS) technique, has been proposed.[7,8] The SLS approach is based on monitoring the changes in laser-generated ultrasound as a laser source is scanned over a defect. Changes in amplitude and frequency content are observed for ultrasound generated by the laser over uniform and defective areas. The SLS technique offers several advantages over conventional techniques. The SLS technique can be used to detect subwavelength surface-breaking cracks. The SLS technique is not as sensitive as pulse–echo techniques to the orientation of the crack with respect to the generating/receiving transducer. Ultrasonic generators such as contact PZT-transducers and near-contact electromagnetic acoustic transducers (EMATs) are not as easily scanned as a laser beam. Noncontact laser scanning can be easily done on specimens such as turbine disks which have complex geometries.[8] Characteristic SLS signatures in terms of amplitude and spectral content variations have been observed as the generating laser scans over uniform and defective areas. These changes are attributed to both near-field scattering and to changes in the laser generation constraints in uncracked and cracked regions. The SLS approach provides enhanced signal-to-noise performance compared to conventional pitch–catch mode of operation. This is because the presence of an anomaly is indicated by an increase in the amplitude of the detected ultrasonic signal rather than by the presence of a weak echo, and by the variations of the ultrasonic frequency when the SLS is in the vicinity of a discontinuity.

The SLS technique has been demonstrated both experimentally[7,8] and via numerical simulations[8,9] using the

[a]Electronic mail: y-sohn@northwestern.edu
[b]Author to whom correspondence should be addressed. Electronic mail: s-krishnaswamy@northwestern.edu

 0001-4966/2004/115(1)/172/10/$20.00

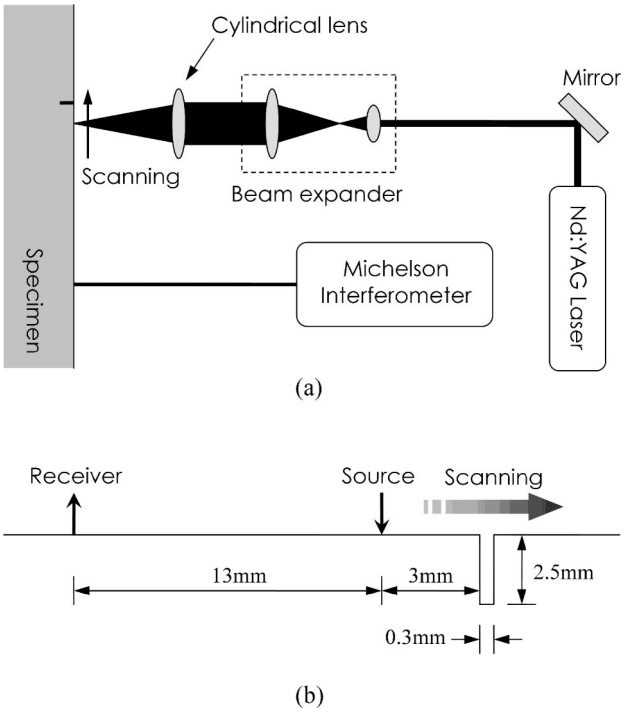


FIG. 1. Experimental setup for the SLLS technique (a) and geometry of test specimen (b).

mass-spring lattice model (MSLM).[10,11] In previous experimental works, the laser source was either a point source or a short line source. In both cases, the thermoelastically generated Rayleigh wave is known to exhibit a bipolar (having both positive and negative going) displacement normal to the surface. In this work, we extend the SLS technique to the case of a true thermoelastic line source, which produces a negative-going monopolar Rayleigh wave displacement.[12–14] We demonstrate that as the scanning laser *line* source (SLLS) approaches a surface-breaking crack, the expected Rayleigh wave arrival signal changes from a monopolar to a bipolar shape and its signal amplitude changes. This dramatic transition, which can be the basis for a novel NDE inspection technique for shallow cracks, is demonstrated both experimentally and via numerical simulations using a modified MSLM technique.

II. SCANNING LASER LINE SOURCE TECHNIQUE—EXPERIMENT

The major components of the experimental apparatus are illustrated schematically in Fig. 1(a). Surface acoustic waves were generated using a Q-switched Nd:YAG laser (Big Sky Laser model CFR 200, 20-Hz pulse rate, 1064-nm wavelength) on an aluminum specimen which was of sufficiently large dimensions that it could be regarded as a half-space. The laser beam was expanded using two spherical lenses and then focused into a line onto the specimen using a cylindrical lens. The line source was approximately measured as 30 mm in length and 200 μm in width (actually the full width at half maximum, FWHM, since the laser beam was nearly Gaussian). The rise time of the laser pulse was 20 ns. The incident laser energy on the specimen was kept low enough (less than 10 mJ) to avoid damaging the aluminum specimen. To keep the apparatus simple for demonstration purposes, the specimen was optically polished, and a quadrature-stabilized Michelson interferometer using a He–Ne (633-nm wavelength) laser was used to detect the normal displacements at the surface. Again, to keep the optics simple, the detection beam was not tightly focused. The size of the detection beam was approximately 600 μm of FWHM assuming Gaussian profile, and this has implications on the bandwidth of detection as explained later in this paper. In real applications on unpolished specimens, a Fabry–Perot or adaptive two-wave mixing interferometers can be used[15,16] instead of the Michelson interferometer, and the beam should be more tightly focused. The aluminum specimen had a surface-breaking EDM notch of 30-mm length, 0.3-mm width, and 2.5-mm depth [Fig. 1(b)]. The SLLS was scanned over a 3.5-mm length in steps of 250 μm across the specimen surface passing over the surface-breaking defect. The interferometer beam was fixed at a distance of 13 mm from the first scanning position. The distance between the first scanning position and the nearest face of the defect was 3 mm. Note that a 30-mm line source is long enough to generate monopolar Rayleigh wave within the source to receiver distance range used in this work (maximum 16.5 mm); that is, the long line source has an extended near field that includes the receiving point.

The left column of Fig. 2 shows the time trace of normal displacements obtained from experiments for various positions of the SLLS. The SLLS position is defined as the distance between the receiver and the center point of the scanning laser source width. (The right column of Fig. 2 is obtained from numerical simulations and will be discussed later.) When the source is far ahead of the defect [Fig. 2(a)] the Rayleigh wave is seen to be monopolar, as expected for a thermoelastic line source. The phase of the reflected Rayleigh wave is not changed by the defect, maintaining the monopolar negative displacement of the incident wave. A significant portion of the incident energy is reflected in this case due to the large depth of the defect. The reflected Rayleigh wave appears to be of slightly lower frequency, and this is attributed to the fact that the defect corner is slightly rounded, and not truly 90°. Other than this, the overall behavior is similar to reflection by a 90° corner.[17] Incident and reflected P waves are not clearly visible in the experiments due to limited detection sensitivity. As the source scans the surface and approaches the defect [Fig. 2(b)] the ultrasonic displacement at the expected Rayleigh wave arrival time undergoes significant changes. There is an increasing positive amplitude signal at the time between the direct and reflected Rayleigh wave arrivals. As the source moves closer to the defect, the signal at the Rayleigh wave arrival time becomes increasingly bipolar, taking its maximum peak-to-peak amplitude when the source is just ahead of the defect [Fig. 2(c)]. Clearly, a bipolar signal cannot be produced by simple interference of the direct and reflected monopolar Rayleigh waves. These dramatic changes in the signal must therefore come from the near-field surface and bulk wave interaction between the source and the defect, including scattering from both the corner and the planar sides of the defect. Furthermore, when the source is close to the defect, a significant

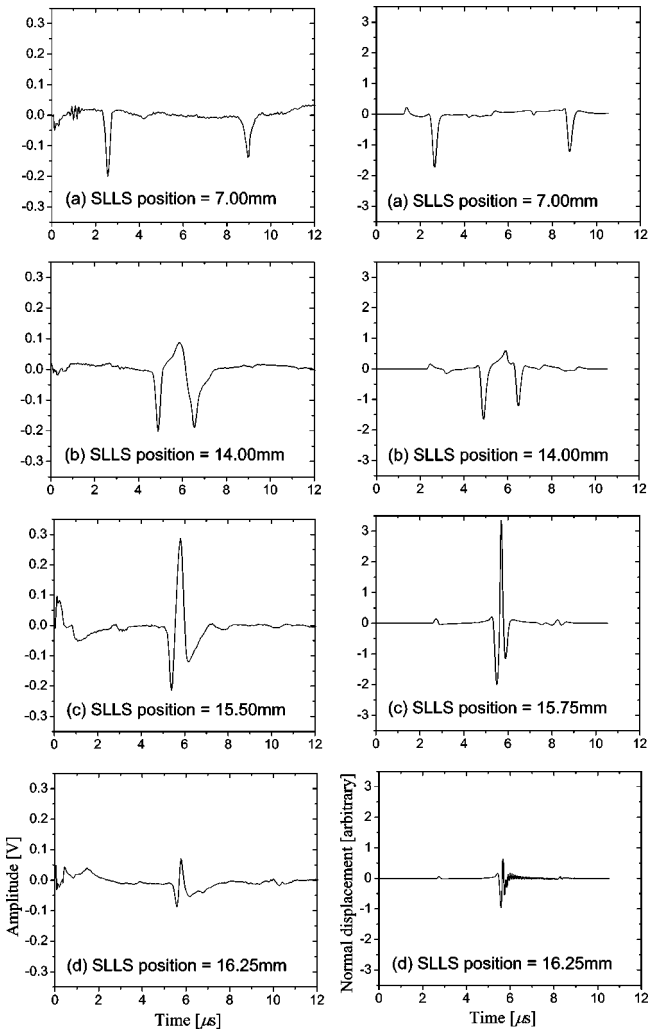


FIG. 2. Normal displacement amplitude for various source positions: experiment (left column) and numerical simulation (right column).

portion of the bulk waves that are generated by the laser is also reflected or scattered by the defect (without spreading away into bulk as would happen if the source and the flaw were farther apart). These can also contribute to the increased signal seen at the receiver. Finally, as the source passes behind the defect [Fig. 2(d)] the amplitude of the Rayleigh wave decreases and almost vanishes. In this case, most of the Rayleigh wave energy is screened because the defect depth is much larger than the wavelength of the Rayleigh wave.

If the peak-to-peak amplitude of the ultrasonic signal within a prescribed window centered around the expected Rayleigh wave arrival time is plotted as a function of the SLLS position, a characteristic SLLS signature shown in Fig. 3 is observed (the numerically simulated results in this figure will be discussed later). The peak-to-peak amplitudes in this figure are normalized by the peak-to-peak value when the source is far ahead of the defect. The defect position is indicated by dotted lines. As the source approaches the defect, the peak-to-peak amplitude increases and reaches its maximum when the source is located just ahead of the defect. As the center of the SLLS passes over the defect, the peak-to-peak amplitude decreases rapidly and goes to almost zero

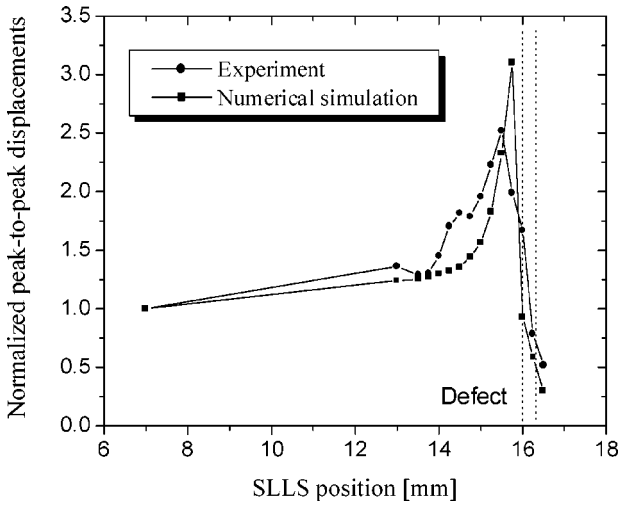


FIG. 3. Peak-to-peak ultrasonic amplitude versus SLLS position.

due to screening by the defect. The peak-to-peak amplitude can be used as a characteristic signature for identifying underlying surface-breaking flaws.

The experiments described above indicate that dramatic changes in the signal at the expected Rayleigh wave arrival window are observed as the SLLS scans over a surface-breaking crack. In the following sections, we will develop a numerical simulation of the SLLS technique to help shed more light on the interaction of the source with the defect which gives rise to the characteristic SLLS signatures observed in the experiments.

III. SCANNING LASER LINE SOURCE TECHNIQUE—NUMERICAL SIMULATION

We will now explore a model to simulate the observed SLLS signatures in order to help understand the SLLS behavior and to aid in optimizing the inspection setup. The model used incorporates two features: (i) the laser ultrasonic source is modeled as a shear-dipole distribution on the surface of a structure, using well-known principles of laser ultrasonics,[2,18] and (ii) the subsequent propagation and scattering of elastodynamic waves in the structure are modeled using a modified mass-spring lattice model.

A. A hybrid MSLM/FDM scheme for simulation of elastodynamic propagation

Of the many numerical methods that allow the simulation and analysis of elastic wave propagation and scattering,[19] we use the mass-spring lattice model (MSLM) in view of its simplicity and reasonable accuracy.[10,11] On traction-free boundaries, however, the MSLM does not always provide highly accurate results.[20] This is because in standard MSLM techniques, the traction-free boundary condition is approximated by simply disconnecting the springs on the simulated boundaries, without actually ensuring that the mathematical expression for the traction-free condition is rigorously imposed. While in many cases involving simulations of bulk waves this has been shown not to be a serious deficiency, this is not the case for SLLS simulations which

inherently depend on surface acoustic wave propagation and scattering from traction-free crack faces. To improve the accuracy of the numerical simulations of the SLLS technique, it is therefore necessary to adopt a more accurate method on the boundary.[20]

Yamawaki *et al.*[21] have suggested a discretized nodal formulation of the wave equation under plane strain conditions in isotropic materials, and this has been shown to provide better results in 2D elastodynamic simulations when traction-free boundaries are involved. Specifically, the nodal equations, which are based on finite difference considerations, enable precise imposition of the mathematical condition of traction-free boundaries, and the scheme (hereafter called ''FDM'') has been shown to provide good results for surface acoustic wave propagation. It is fortunate that the FDM method can readily be incorporated at the traction-free boundaries of an MSLM grid, since both models have been developed under the same wave equation (2D plane strain elasticity), and involve using the same nodes which represent the centers of mass of discretized meshed areas. Therefore, the FDM can be directly adopted on planar-, concave-, and convex-shaped traction-free surfaces of an MSLM grid. A schematic of this hybrid MSLM/FDM model is shown in Fig. 4. The discretized equations of motion in this case are given below for interior nodes (node A, using standard MSLM formulation), for nodes on a traction-free planar surface (node B), at a traction-free convex corner (node C), and at a traction-free concave corner (node D)

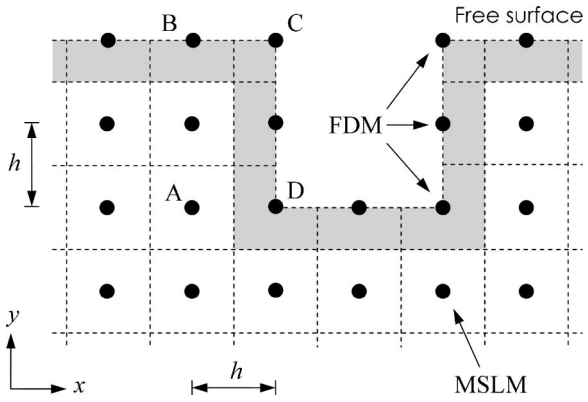


FIG. 4. Schematic of hybrid MSLM/FDM model.

NodeA:

$$\rho(u_{i,j}^{k+1}+u_{i,j}^{k-1}-2u_{i,j}^{k})/(\Delta t)^2$$
$$=F_x+k_1(u_{i+1,j}^{k}+u_{i-1,j}^{k}-2u_{i,j}^{k})/h^2+k_2(u_{i+1,j+1}^{k}+u_{i+1,j-1}^{k}+u_{i-1,j+1}^{k}+u_{i-1,j-1}^{k}-4u_{i,j}^{k})/2h^2+k_2\beta(u_{i+1,j+1}^{k}$$
$$+u_{i+1,j-1}^{k}+u_{i-1,j+1}^{k}+u_{i-1,j-1}^{k}-4u_{i,j}^{k})/2h^2+k_2(v_{i+1,j+1}^{k}-v_{i+1,j-1}^{k}-v_{i-1,j+1}^{k}+v_{i-1,j-1}^{k})/2h^2+k_2\beta$$
$$\times(-v_{i+1,j+1}^{k}+v_{i+1,j-1}^{k}+v_{i-1,j+1}^{k}-v_{i-1,j-1}^{k})/2h^2, \tag{1}$$

$$\rho(v_{i,j}^{k+1}+v_{i,j}^{k-1}-2v_{i,j}^{k})/(\Delta t)^2$$
$$=F_y+k_3(v_{i,j+1}^{k}+v_{i,j-1}^{k}-2v_{i,j}^{k})/h^2+k_2(v_{i+1,j+1}^{k}+v_{i+1,j-1}^{k}+v_{i-1,j+1}^{k}+v_{i-1,j-1}^{k}-4v_{i,j}^{k})/2h^2$$
$$+k_2\beta(v_{i+1,j+1}^{k}+v_{i+1,j-1}^{k}+v_{i-1,j+1}^{k}+v_{i-1,j-1}^{k}-4v_{i,j}^{k})/2h^2+k_2(u_{i+1,j+1}^{k}-u_{i+1,j-1}^{k}-u_{i-1,j+1}^{k}$$
$$+u_{i-1,j-1}^{k})/2h^2+k_2\beta(-u_{i+1,j+1}^{k}+u_{i+1,j-1}^{k}+u_{i-1,j+1}^{k}-u_{i-1,j-1}^{k})/2h^2. \tag{2}$$

Node B:

$$\rho(u_{i,j}^{k+1}+u_{i,j}^{k-1}-2u_{i,j}^{k})/(\Delta t)^2$$
$$=F_x+C_{11}(2u_{i+1,j-1}^{k}-4u_{i,j-1}^{k}+2u_{i-1,j-1}^{k}+6u_{i+1,j}^{k}-12u_{i,j}^{k}+6u_{i-1,j}^{k})/8h^2+C_{12}(-4v_{i+1,j-1}^{k}$$
$$+4v_{i-1,j-1}^{k}+4v_{i+1,j}^{k}-4v_{i-1,j}^{k})/8h^2+C_{33}(u_{i+1,j-1}^{k}+6u_{i,j-1}^{k}+u_{i-1,j-1}^{k}-u_{i+1,j}^{k}-6u_{i,j}^{k}$$
$$-u_{i-1,j}^{k})/8h^2+C_{33}(-3v_{i+1,j-1}^{k}+3v_{i-1,j-1}^{k}-v_{i+1,j}^{k}+v_{i-1,j}^{k})/8h^2, \tag{3}$$

$$\rho(v_{i,j}^{k+1}+v_{i,j}^{k-1}-2v_{i,j}^{k})/(\Delta t)^2$$
$$=F_y+C_{11}(v_{i+1,j-1}^{k}+6v_{i,j-1}^{k}+v_{i-1,j-1}^{k}-v_{i+1,j}^{k}-6v_{i,j}^{k}-v_{i-1,j}^{k})/8h^2+C_{12}(-3u_{i+1,j-1}^{k}+3u_{i-1,j-1}^{k}$$
$$-u_{i+1,j}^{k}+v_{i-1,j}^{k})/8h^2+C_{33}(2v_{i+1,j-1}^{k}-4v_{i,j-1}^{k}+2v_{i-1,j-1}^{k}+6v_{i+1,j}^{k}-12v_{i,j}^{k}+6v_{i-1,j}^{k})/8h^2$$
$$+C_{33}(-4u_{i+1,j-1}^{k}+4u_{i-1,j-1}^{k}+4u_{i+1,j}^{k}-4u_{i-1,j}^{k})/8h^2. \tag{4}$$

Node C:

$$\rho(u_{i,j}^{k+1}+u_{i,j}^{k-1}-2u_{i,j}^{k})/(\Delta t)^2$$
$$=F_x+C_{11}(u_{i-1,j-1}^{k}+3u_{i-1,j}^{k}-u_{i,j-1}^{k}-3u_{i,j}^{k})/4h^2+C_{12}(3v_{i-1,j-1}^{k}-3v_{i-1,j}^{k}+v_{i,j-1}^{k}-v_{i,j}^{k})/4h^2$$
$$+C_{33}(u_{i-1,j-1}^{k}-u_{i-1,j}^{k}+3u_{i,j-1}^{k}-3u_{i,j}^{k})/4h^2+C_{33}(3v_{i-1,j-1}^{k}+v_{i-1,j}^{k}-3v_{i,j-1}^{k}-v_{i,j}^{k})/4h^2, \quad (5)$$

$$\rho(v_{i,j}^{k+1}+v_{i,j}^{k-1}-2v_{i,j}^{k})/(\Delta t)^2$$
$$=F_y+C_{11}(v_{i-1,j-1}^{k}-v_{i-1,j}^{k}+3v_{i,j-1}^{k}-3v_{i,j}^{k})/4h^2+C_{12}(3u_{i-1,j-1}^{k}+u_{i-1,j}^{k}-3u_{i,j-1}^{k}-u_{i,j}^{k})/4h^2$$
$$+C_{33}(v_{i-1,j-1}^{k}+3v_{i-1,j}^{k}-v_{i,j-1}^{k}-3v_{i,j}^{k})/4h^2+C_{33}(3u_{i-1,j-1}^{k}-3u_{i-1,j}^{k}+u_{i,j-1}^{k}-u_{i,j}^{k})/4h^2. \quad (6)$$

Node D:

$$\rho(u_{i,j}^{k+1}+u_{i,j}^{k-1}-2u_{i,j}^{k})/(\Delta t)^2$$
$$=F_x+C_{11}(2u_{i-1,j-1}^{k}+9u_{i-1,j}^{k}+u_{i-1,j+1}^{k}-4u_{i,j-1}^{k}-15u_{i,j}^{k}-u_{i,j+1}^{k}+2u_{i+1,j-1}^{k}+6u_{i+1,j}^{k})/12h^2$$
$$+C_{12}(4v_{i-1,j-1}^{k}-v_{i-1,j}^{k}-3v_{i-1,j+1}^{k}+v_{i,j}^{k}-v_{i,j+1}^{k}-4v_{i+1,j-1}^{k}+4v_{i+1,j}^{k})/12h^2$$
$$+C_{33}(2u_{i-1,j-1}^{k}-4u_{i-1,j}^{k}+2u_{i-1,j+1}^{k}+9u_{i,j-1}^{k}-15u_{i,j}^{k}+6u_{i,j+1}^{k}+u_{i+1,j-1}^{k}-u_{i+1,j}^{k})/12h^2$$
$$+C_{33}(4v_{i-1,j-1}^{k}-4v_{i-1,j+1}^{k}-v_{i,j-1}^{k}+v_{i,j}^{k}+4v_{i,j+!}^{k}-3v_{i+1,j-1}^{k}-v_{i+1,j}^{k})/12h^2, \quad (7)$$

$$\rho(v_{i,j}^{k+1}+v_{i,j}^{k-1}-2v_{i,j}^{k})/(\Delta t)^2$$
$$=F_y+C_{11}(2v_{i-1,j-1}^{k}-4v_{i-1,j}^{k}+2v_{i-1,j+1}^{k}+9v_{i,j-1}^{k}-15u_{i,j}^{k}+6v_{i,j+1}^{k}+v_{i+1,j-1}^{k}-v_{i+1,j}^{k})/12h^2$$
$$+C_{12}(4u_{i-1,j-1}^{k}-4u_{i-1,j+1}^{k}-u_{i,j-1}^{k}+u_{i,j}^{k}+4u_{i,j+1}^{k}-3u_{i+1,j-1}^{k}-u_{i+1,j}^{k})/12h^2+C_{33}(2v_{i-1,j-1}^{k}+9v_{i-1,j}^{k}$$
$$+v_{i-1,j+1}^{k}-4v_{i,j-1}^{k}-15v_{i,j}^{k}-v_{i,j+1}^{k}+2v_{i+1,j-1}^{k}+6v_{i+1,j}^{k})/12h^2+C_{33}(4u_{i-1,j-1}^{k}-u_{i-1,j}^{k}$$
$$-3u_{i-1,j+1}^{k}+u_{i,j}^{k}-u_{i,j+1}^{k}-4u_{i+1,j-1}^{k}+4u_{i+1,j}^{k})/12h^2. \quad (8)$$

Here, h and Δt are grid size and time step, respectively. $u_{i,j}^{k+1}$, $v_{i,j}^{k+1}$ denote the x- and y-direction displacements of the node at the position (i,j) and at the time $t=k\Delta t$. F_x and F_y are the body force exerted on the node in the x and y directions, respectively. ρ is mass density, and C_{11}, C_{12}, and C_{33} are stiffness coefficients of material. k_1, k_2, k_3, α, and β are constants[10] related to elastic property of material represented by stiffness coefficients C_{11}, C_{12}, and C_{33}.

We first demonstrate that the hybrid model performs better than standard MSLM calculations in predicting surface acoustic waves (Rayleigh waves). An aluminum ($C_{11}=9.1\times10^{10}$ N/m^2, $C_{12}=4.1\times10^{10}$ N/m^2, $C_{33}=(C_{11}-C_{12})/2$, $\rho=2600$ kg/m^3) block was chosen as the test specimen. The grid size chosen was $h=20$ μm and a corresponding time step chosen was $\Delta t=1.69$ ns. For consistency of modeling, the same material properties, grid size, and time steps are used in all the calculations reported in this paper. The grid size is chosen so that sufficiently small flaws can be modeled accurately in the SLLS simulations reported later. To investigate Rayleigh wave propagation, a 9-MHz one-cycle sinusoidal wave was generated on the free surface of the standard MSLM and the hybrid MSLM/FDM models. The displacements of the Rayleigh wave were measured at two points on the surface and the resulting Rayleigh wave speeds were calculated using cross correlation. The analytically calculated Rayleigh wave speed based on the material properties is 2880 m/s. In the MSLM model, the calculated Rayleigh wave speed was 2665 m/s (7.5% error), and in the hybrid model the speed was calculated to be 2838 m/s (1.5% error). To analyze the Rayleigh wave motions near the free surface, the tangential and normal displacements of nodes at different depths are plotted in Fig. 5. The displacements and depths

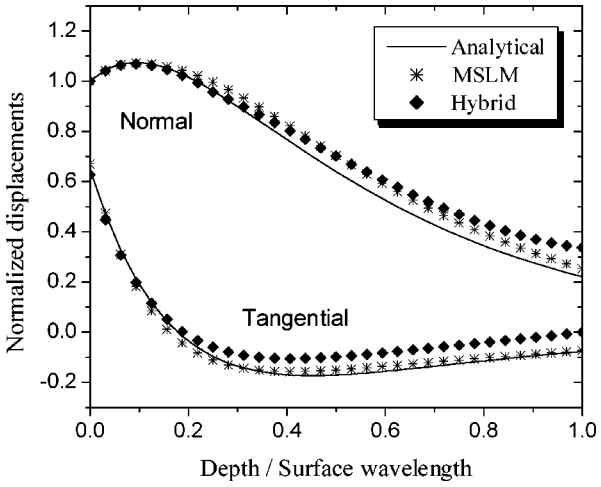


FIG. 5. Calculated Rayleigh wave particle motion near the free surface.

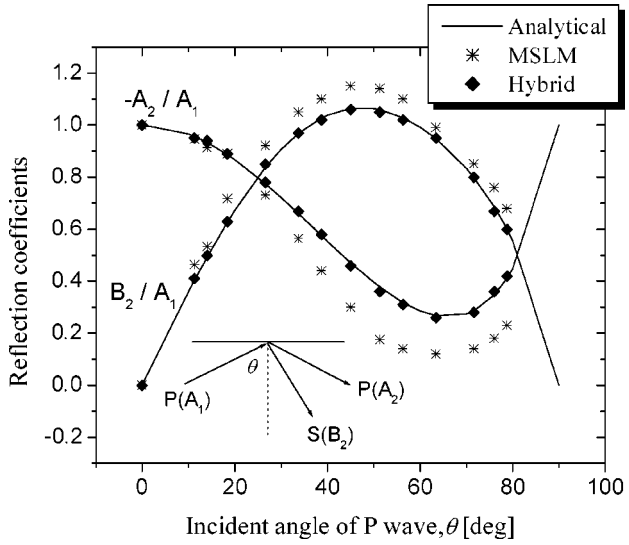


FIG. 6. Reflection coefficients for various incident angles of P waves.

are normalized by the maximum normal displacement and wavelength of Rayleigh wave, respectively. Measured displacements are obtained from the displacements at the nodal points of the dash-lined area on the surface (gray areas in Fig. 4). That is, the calculated values are actually the displacements of the center of the meshed surface area. Both the MSLM and hybrid model show good agreement with each other and are consistent with analytical results,[22] especially near the surface where the greater portion of the Rayleigh wave energy propagates. The deviation from the analytical distribution at greater depths is mitigated by the fact that the actual displacements are vanishingly small in this region.

We next demonstrate that the hybrid model performs better than standard MSLM calculations in predicting reflections of bulk waves from traction-free boundaries. It is necessary to establish quantitative accuracy of reflections at traction-free surfaces because such reflections are a significant contributing factor in the SLLS technique. Following the lines of the Rayleigh wave simulations, reflection coefficients of directly reflected or mode converted waves by incident harmonic plane P (longitudinal) and S (shear) wave (9 MHz, one cycle) were investigated using both the standard MSLM and the MSLM/FDM hybrid models. Numerical and analytical results with respect to various incident angles θ are shown in Fig. 6 and Fig. 7. A_1, B_1, A_2, and B_2 represent the amplitude of incident P, incident S, reflected P, and reflected S waves, respectively. Though both numerical results show reasonably good agreement with analytical results[22] in general, it should be noted that the reflection coefficients calculated by the standard MSLM shows significant error in the case of reflected P wave by incident P wave when the incident angle is larger than 30°. The hybrid MSLM/FDM method predicts quantitatively better results.

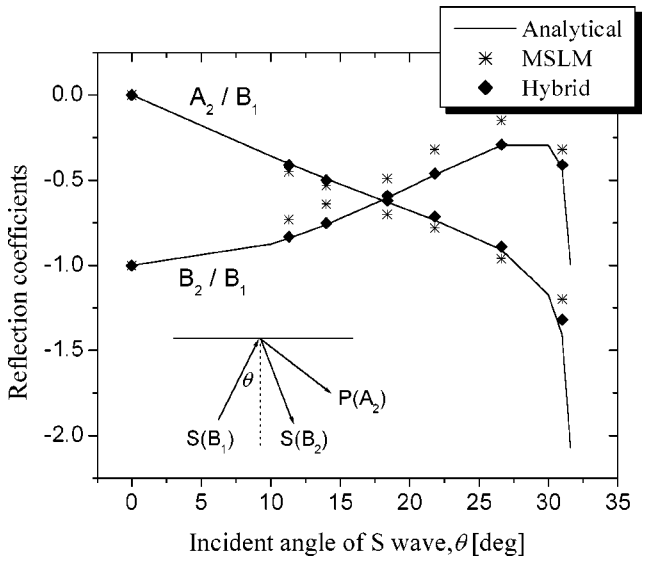


FIG. 7. Reflection coefficients for various incident angles of S waves.

Considering the results above, we expect that the hybrid method can provide more accurate results when simulating the SLLS technique than the standard MSLM technique. We therefore will use the hybrid MSLM/FDM method for the numerical simulation of the SLLS technique.

B. Modeling of a thermoelastic laser line source

The physical mechanism of laser generation of ultrasound is well established.[2] When a laser pulse impinges a material surface, part of the electromagnetic energy of the laser source is absorbed by the material and is converted into heat. The rapid heat absorption in a small, localized region of the object leads to rapid thermal expansion in this region, and this in turn generates ultrasonic waves into the bulk of the material. We will restrict attention to the case where the laser-induced temperature rise is within the thermoelastic regime without producing ablation or plasma formation of the material. A number of simplifications can be made to model thermoelastic generation. First, we assume that no heat is lost from the object by thermal conduction or radiation (a reasonable assumption over the time scale of interest in modeling). Furthermore, it is considered that no significant optical penetration occurs below the surface of the test material since the absorption depth for 1064-nm laser wavelength on metallic materials such as aluminum or titanium (on which the laser ultrasonic test is most commonly applied) is in the nanometer range.

It is well established that a thermoelastic source at a point in the interior of an isotropic material creates a volumetric source of expansion, the elastodynamics of which can be modeled using three orthogonal force dipoles.[2] In the SLLS technique of interest, the thermoelastic laser source is not buried within the material, but rather is applied on the surface. In this case, it has been shown that the source becomes a surface center of expansion, which results in a pair of orthogonal dipoles parallel to the surface.[23] For the case of an infinitesimally thin line laser source impinging on a surface (leading to a plane strain formulation), the surface center of expansion essentially becomes a single shear dipole parallel to the surface. For spatially distributed laser sources, the model can be extended to include a spatially distributed set of dipoles.

As the dipole strengths are proportional to the laser-induced temperature variation, the spatial and temporal temperature distribution across the surface should be considered.

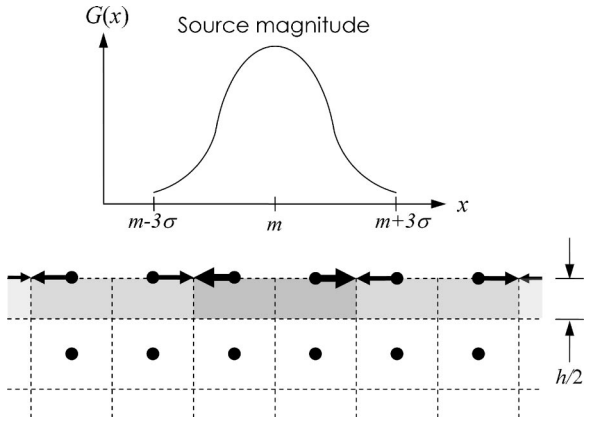


FIG. 8. Discretized laser source (dipole model) showing Gaussian spatial distribution.

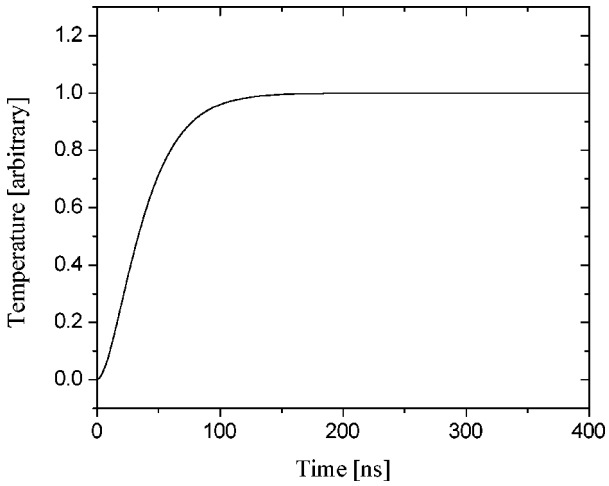


FIG. 9. Temporal profile of the temperature variation at the surface.

For the case of a broad laser line source, the dipole magnitude $D(x,t)$ at location x and time t is given by

$$D(x,t)=AT(x,t). \tag{9}$$

Here, A is a constant factor that includes the efficiency of conversion of electomagnetic to thermal energy, along with elastic and thermal properties of the material, and $T(x,t)$ is the spatial and temporal distribution of the temperature rise due to absorption of the laser energy.

The temperature profile can be calculated knowing the heat source and assuming adiabatic conditions. The time derivative of the temperature is proportional to the heat source, which in turn is proportional to the laser pulse, and is given by[23]

$$\dot{T}(x,t)=QG(x)q(t), \tag{10}$$

where Q is a constant factor related to the thermal properties of material, $G(x)$ is the spatial distribution, and $q(t)$ is the temporal pulse shape of the laser source. The laser output is assumed to have a Gaussian spatial distribution of the form

$$G(x)=e^{-(x-m)^2/2\sigma^2}, \tag{11}$$

where σ is standard deviation and m is mean value of the distribution. Numerically, the source is actually truncated such that the end positions of the finite width are $x=m\pm 3\sigma$, at which points the value of $G(x)$ is sufficiently small so as not to cause serious numerical discontinuity. The temporal pulse shape for a Q-switched laser can be approximated by[24]

$$q(t)=(t/\tau)e^{(1-t/\tau)}, \tag{12}$$

where τ is the pulse rising time of the laser source. The resulting temperature profile is therefore

$$T(x,t)=QG(x)h(t), \tag{13}$$

where

$$h(t)=\tau(1-e^{-t/\tau})-te^{-t/\tau}. \tag{14}$$

Note that optical reflectivity and material, thermal, and elastic properties are absorbed into the constant $A\cdot Q$ for the dipole magnitudes, for convenience. The spatial distribution of shear dipoles given by Eq. (11) is applied in its discretized form in the MSLM as shown in Fig. 8. The normalized temporal profile of the temperature rise ($\tau=20$ ns) is shown in Fig. 9.

It should be noted that in an earlier work on the SLS technique,[9] the temporal distribution for the temperature profile was assumed to be similar to that of the incident laser pulse. That is, the temporal profile of the temperature was assumed to increase and then decrease with time, thereby inherently assuming rapid thermal diffusion. Such a temporal profile enabled the simulation of the bipolar nature of the Rayleigh wave produced by such point or short line sources. The model presented here is more appropriate for simulating the monopolar Rayleigh wave pulse under thermoelastic generation using a laser line source in the absence of thermal diffusion.

Figure 10 shows the calculated normal and tangential displacements on the surface at a location 2.25 mm to the right of the excitation laser line source (of FWHM 200 μm). The displacements are normalized by the negative maximum value of normal displacement. The normal displacements of the Rayleigh wave show the expected negative monopolar shape instead of the bipolar waveform associated with a

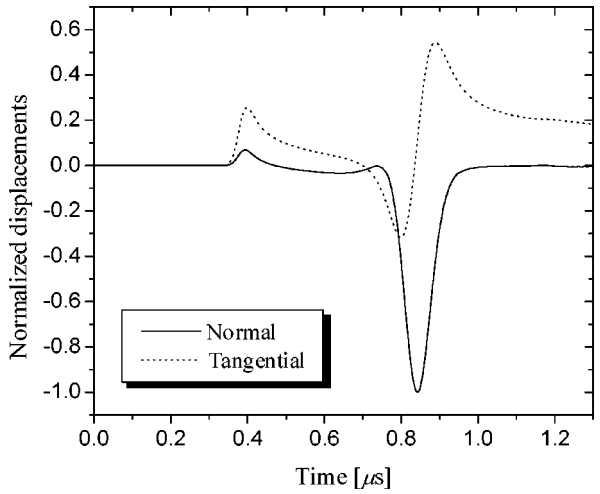


FIG. 10. Numerically calculated displacements at a distance of 2.25 mm from the source.

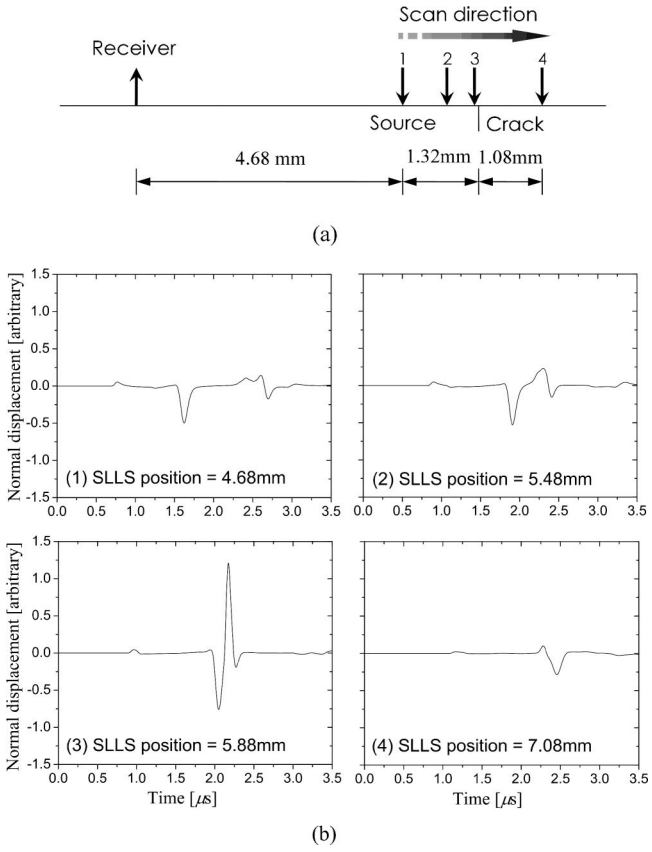


FIG. 11. Schematic of the SLLS technique (a) and normal surface displacements for selected SLLS positions (b).

point or relatively short line source. It is observed that the numerically calculated displacements are in good agreement with established analytical and experimental results.[12–14] The measured Rayleigh wave pulse width (FWHM) is around 80 ns, and the spatial FWHM width of the Rayleigh wave pulse is about 300 μm.

C. Modeling of SLLS technique

Having established that the hybrid MSLM/FDM method in conjunction with the shear dipole distribution model predicts the monopolar Rayleigh wave with reasonable accuracy, we turn to the simulation of the SLLS technique. A scanning laser line source with the same spatial and temporal characteristics as described in subsection B was modeled on an aluminum half-space containing a surface-breaking crack. The schematic of the SLS simulation is shown in Fig. 11(a). The receiver is fixed at distance of 6 mm from the crack and the laser source is scanned over the surface of the block in steps of 80 μm. The depth of the surface-breaking crack is 180 μm, which is smaller than the spatial FWHM of the incident Rayleigh wave (300 μm). The crack width is 20 μm (equal to the grid space, h) which is the minimum width that can be modeled using this discretization. The time traces of the normal displacement for various source positions are shown in Fig. 11(b). The direct and the crack-reflected Rayleigh waves can be seen when the source position is far ahead of the crack (position 1). As the laser source approaches close to the crack (position 2), the direct and the reflected waves get closer together. When the source is just

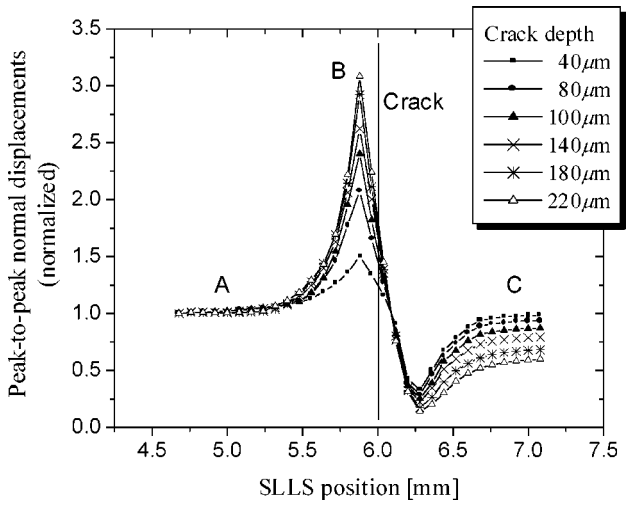


FIG. 12. Peak-to-peak displacements versus SLLS position for various crack depths.

ahead of the crack (position 3), significant changes in the signal occur, with the monopolar Rayleigh wave displacement giving place to a bipolar displacement. As the source passes behind the crack (position 4), the amplitude of the Rayleigh wave signal decreases due to screening by the crack. It is seen that the simulation results are consistent with those observed experimentally, indicating that the primary mechanisms involved in the SLLS signature change are captured by the simulation. While the precise mechanism remains to be determined, we can infer that the shape and amplitude changes are due to the surface and bulk waves generated by the laser source interacting with the near-field defect, resulting in a part of the bulk waves that would have normally been radiated into the solid being scattered back into the direction of the receiver.

Similar simulations were done for various crack depths ranging from 40 to 220 μm. The selected crack depths are much smaller than the spatial FWHM of the incident Rayleigh wave (which can be thought of as related to the spatial wavelength of the Rayleigh wave). A plot of the peak-to-peak amplitude of the calculated normal displacements versus the SLLS position for various crack depths is shown Fig. 12. After a region of constant amplitude (region A, far from the crack), the peak-to-peak normal displacements increase significantly as the source approaches the crack and reaches its maximum value when the center of the source is just ahead of the crack (region B). As the source passes over the crack the displacements decrease noticeably and become stable when the source is far behind the crack (region C). In the case of the smallest crack depth shown, the difference in amplitude between region A and C is not very large because the greater portion of the incident energy transmits past the crack. However, even in this case, the characteristic SLLS signature is visible.

The changes in the peak-to-peak displacements near or behind the crack can be used to gauge crack depths. Figure 13 shows the peak-to-peak amplitude changes versus various crack depths for two SLLS positions: when the source is just ahead of the crack, and far behind the crack. The amplitudes are normalized with the peak-to-peak displacement of the

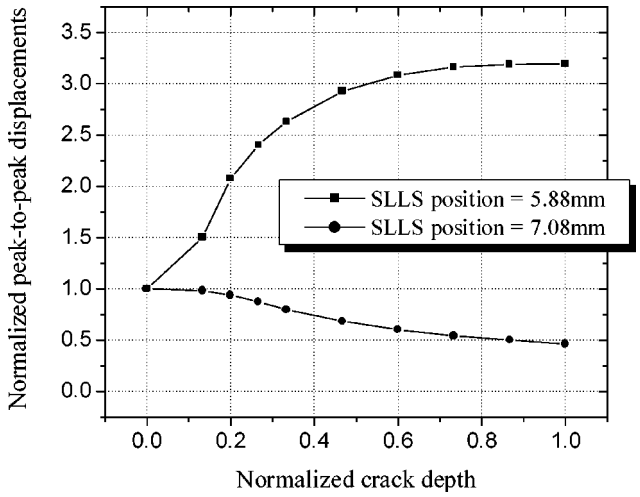


FIG. 13. Normalized normal displacement versus normalized crack depth at two specific SLLS positions.

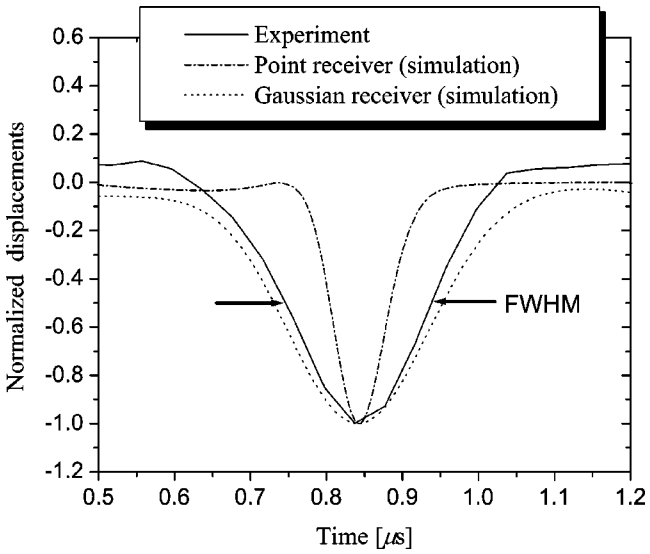


FIG. 14. Normal displacements of measured and simulated Rayleigh waves.

Rayleigh wave obtained in the absence of a crack. The crack depth is normalized with spatial FWHM of incident Rayleigh wave (300 μm). The decrease in the amplitude of the signal measured behind the crack indicates crack shielding and is the conventional transmit characteristics signature relied upon in the pitch–catch method. The increasing peak-to-peak amplitude signature of the signal measured ahead of the crack clearly shows the advantage of the SLLS technique with respect to the pitch–catch method in that (i) the variations are larger, and (ii) an increase in signal is more readily monitored with enhanced signal-to-noise ratio than a decrease in signal amplitude.

Finally, the hybrid MSLM/FDM method is used to simulate the SLLS experiments reported in Sec. II with the same geometric features of the specimen and same characteristics of the scanning laser source and receiver. The grid space used was 20 μm and the corresponding time step was 1.69 ns in aluminum. A line source of 200-μm FWHM and 20-ns rising time was modeled in the same manner as above to provide a monopolar Rayleigh wave. The receiver was modeled as a Gaussian beam of 600-μm FWHM. The Rayleigh wave displacements computed at a single point will obviously be different from those measured in the experiments due to the spatial averaging over the detection beamwidth. This is clearly seen in Fig. 14, which shows the computed normal displacements and the measured amplitudes (normalized by their respective maximum negative values). The Rayleigh wave pulse width computed at a single point is about 80 ns, which is much smaller than the experimentally measured value of about 200 ns. Using a weighted average of the computed displacements over the detection beamwidth (essentially a convolution of the computed displacements over the Gaussian beamwidth of the detection beam), the measured and the computed Rayleigh waves are seen to be in much better agreement. The numerical simulation results presented below are all convolved with the detection beamwidth to facilitate comparison between the experimental results and the model.

Figure 2 shows the time trace of the surface wave amplitude obtained from the experiments (left column) and the numerical simulation (right column) of the SLLS for various source positions. The agreement between the experiment and the model is seen to be reasonably good. The SLLS peak-to-peak amplitude signatures are shown in Fig. 3. The experimental and the simulation results show slight difference in the value of the maximum and its position. This is possibly due to the fact that the experimental defect was slightly rounded (and therefore was slightly wider). In all, the agreement between the experiments and the model is reasonable, suggesting that the primary mechanisms responsible for the SLLS signatures are well captured by the model.

IV. CONCLUSION

The salient points of this paper are

(1) Experiments using a scanning pulsed laser line source on an aluminum block containing an EDM notch indicate a change in the Rayleigh wave normal displacement from a negative monopolar one to a bipolar one, with an increase in the peak-to-peak amplitude as the SLLS approaches the defect.
(2) Thermoelastic generation of Rayleigh waves in an isotropic material by a scanning laser line source has been modeled numerically using a hybrid MSLM/FDM technique. The laser line source is modeled as a distribution of shear dipoles.
(3) The model predicts the same characteristic monopolar to bipolar Rayleigh wave signal transition as the SLLS approaches the crack.
(4) The agreement between the model and the experimental results suggests that the primary mechanisms responsible for the observed signal variations are well captured by the model. While the precise mechanism is yet to be determined, the shape and amplitude changes are due to the surface and bulk waves generated by the laser source interacting with the near-field defect, resulting in a part of the bulk waves that would have normally been radiated into the solid being scattered back into the direction of the receiver.

ACKNOWLEDGMENTS

This work was supported by the FAA through Grant No. DTFA03-98-F-IA029.

[1] J. Krautkrämer and H. Krautkrämer, *Ultrasonic Testing of Materials*, 4th ed. (Springer, New York, 1990).
[2] C. B. Scruby and L. E. Drain, *Laser Ultrasonics: Techniques and Applications* (Hilger, New York, 1990).
[3] J. A. Cooper, R. A. Crosbie, R. J. Dewhurst, A. D. W. McKie, and S. B. Palmer, ''Surface acoustic wave interactions with cracks and slots: A noncontacting study using lasers,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **33**(5), 462–470 (1986).
[4] Q. Shan and R. J. Dewhurst, ''Surface-breaking fatigue crack detection using laser ultrasound,'' Appl. Phys. Lett. **62**(21), 2649–2651 (1993).
[5] C. M. Scala and S. J. Bowles, ''Laser ultrasonics for surface-crack depth measurement using transmitted near-field Rayleigh waves,'' *Review of Progress in QNDE* (AIP, Melville, NY, 2000), Vol. 19, pp. 327–334.
[6] J. R. Bernstein and J. B. Spicer, ''Hybrid laser/broadband EMAT ultrasonic system for characterizing cracks in metals,'' J. Acoust. Soc. Am. **111**(4), 1685–1691 (2002).
[7] A. K. Kromine, P. A. Fomitchov, S. Krishnaswamy, and J. D. Achenbach, ''Laser ultrasonic detection of surface-breaking discontinuities: Scanning laser source technique,'' Mater. Eval. **58**(2), 173–177 (2000).
[8] P. A. Fomitchov, A. K. Kromine, Y. Sohn, S. Krishnaswamy, and J. D. Achenbach, ''Ultrasonic imaging of small surface-breaking defects using scanning laser source technique,'' *Review of Progress in QNDE* (AIP, Melville, NY, 2002), Vol. 21A, pp. 356–362.
[9] Y. Sohn and S. Krishnaswamy, ''Mass spring lattice modeling of the scanning laser source technique,'' Ultrasonics **39**, 543–551 (2002).
[10] H. Yim and Y. Sohn, ''Numerical simulation and visualization of elastic waves using mass-spring lattice model,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**(3), 549–558 (2000).
[11] H. Yim and Y. Choi, ''Simulation of ultrasonic waves in various types of elastic media using the mass spring lattice model,'' Mater. Eval. **58**(7), 889–896 (2000).
[12] P. A. Doyle and C. M. Scala, ''Near-field ultrasonic Rayleigh waves from a laser line source,'' Ultrasonics **34**, 1–8 (1996).
[13] D. Royer and C. Chenu, ''Experimental and theoretical waveforms of Rayleigh waves generated by a thermoplastic laser line source,'' Ultrasonics **38**, 891–895 (2000).
[14] C. Edwards, T. Stratoudaki, S. Dixon, and S. B. Palmer, ''Laser generated Rayleigh and Lamb waves,'' *Review of Progress in QNDE* (AIP, Melville, NY, 2002), Vol. 21, pp. 284–291.
[15] J.-P. Monchalin, ''Optical detection of ultrasound,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **33**(5), 485–499 (1986).
[16] R. K. Ing and J.-P. Monchalin, ''Broadband optical detection of ultrasound by two-wave mixing in a photorefractive crystal,'' Appl. Phys. Lett. **59**, 3233–3235 (1991).
[17] L. J. Bond, ''A computer model of the interaction of acoustic surface waves with discontinuities,'' Ultrasonics **17**, 71–77 (1979).
[18] J. R. Bernstein and J. B. Spicer, ''Line source representation for laser-generated ultrasound in aluminum,'' J. Acoust. Soc. Am. **107**(3), 1352–1357 (2000).
[19] L. J. Bond, ''Numerical techniques and their use to study wave propagation and scattering: A review,'' in *Elastic Waves and Ultrasonic Nondestructive Evaluation*, edited by S. K. Datta, J. D. Achenbach, and Y. S. Rajapakse (North-Holland, Amsterdam, 1990), pp. 17–27.
[20] H. Yim and C.-J. Lee, ''Quantitative accuracy of the mass-spring lattice model in simulating ultrasonic waves,'' *Review of Progress in QNDE* (AIP, Melville, NY 2002), Vol. 21, pp. 152–156.
[21] H. Yamawaki and T. Saito, ''Numerical calculation of surface waves using new nodal equation,'' Nondestr. Test. Eval. **8–9**, 379–389 (1992).
[22] J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam, 1980).
[23] L. R. F. Rose, ''Point-source representation for laser-generated ultrasound,'' J. Acoust. Soc. Am. **75**(3), 723–732 (1984).
[24] J. F. Ready, *Effects of High-Power Laser Radiation* (Academic, New York, 1971).

Optimization of the performance of the sandwich piezoelectric ultrasonic transducer

Lin Shuyu[a)]
Institute of Applied Acoustics, Shaanxi Normal University, Xian, Shaanxi, 710062, People's Republic of China

(Received 24 January 2002; revised 5 September 2003; accepted 3 November 2003)

The resonance and antiresonance frequency, the effective electromechanical coupling coefficient, and the mechanical quality factor of a sandwich piezoelectric ultrasonic transducer are studied and optimized. The effect of the thickness of thick piezoelectric element electrodes on the transducer performance is analyzed. The effect of the length and position of the piezoelectric elements in the transducer is also studied. It is shown that, although using thick electrodes is beneficial for releasing heat produced by the piezoelectric elements, the effective electromechanical coupling coefficient and the mechanical quality factor are reduced. The length and the position of the piezoelectric elements affect the performances of the transducer. Increasing the length of the piezoelectric elements decreases the mechanical quality factor, but the effective electromechanical coupling coefficient increases. When the length reaches a certain value, the effective electromechanical coupling coefficient reaches a maximum value. When the piezoelectric elements are located at the geometrical center or the displacement node, the effective electromechanical coupling coefficient and the mechanical quality factor are maximized.
[DOI: 10.1121/1.1635836]

PACS numbers: 43.38.Fx, 43.38.Ar [AJZ] Pages: 182–186

I. INTRODUCTION

In high-power ultrasonics and underwater acoustics, the sandwich piezoelectric transducer is widely used. This type of transducers is composed of piezoelectric ceramic elements, sandwiched between two metal blocks. The advantages of this type of transducers are the following: (1) The longitudinal piezoelectric effect is used, so it has a high electromechanical conversion coefficient. (2) Through the use of the metal blocks, the transducer can be operated at a low resonance frequency. (3) Although the piezoelectric ceramic elements are fragile under large tensile stress, it can be operated at high power by bolting the piezoelectric elements between the blocks, placing the piezoelectric material under compressive stress. In this way, a tensile fracture can be avoided. (4) Since piezoelectric ceramic elements are insulators, the elements heat during use at high-power levels, reducing the electromechanical conversion efficiency. The increased loss with increasing temperature can result in a runaway condition, in some cases resulting in the destruction of the transducer. Because of their good heat conduction and contact with the outside environment the metal blocks help dissipate the heat. By choosing the thickness, the cross-sectional area, and the metal material, the temperature coefficient of the piezoelectric elements can be compensated, and the total temperature coefficient of the sandwich transducer can be improved. (5) For the sandwich transducer, by changing the thickness and the cross-sectional dimensions of the piezoelectric elements, the cross-sectional shape and dimensions of the metal blocks, the transducer performance can be optimized.

[a)]Electronic mail: sxsdsxs@snnu.edu.cn

The analysis of the sandwich piezoelectric ultrasonic transducer has been studied for many years and the design theory has been well developed.[1–5] In high-power ultrasonic technology, the electroacoustical efficiency of the transducer is an important parameter. In order to increase the electroacoustical efficiency, a number of factors should be considered. These include the electrical and acoustical matching and the dielectric and mechanical loss. In a high-power ultrasonic transducer, the piezoelectric elements are the main heat source. In order to decrease the heat, thick electrodes between the piezoelectric elements can be used; the thick electrodes can help dissipate the heat. In the following analysis, the effect of thick electrodes on the transducer performance is studied. The inner friction of the material and the contacting interfaces between the metal blocks and the piezoelectric elements produces mechanical losses. For a high-power ultrasonic transducer, the mechanical loss has an important effect on the transducer performance. In previous studies, the effect of the mechanical loss on the transducer performance is seldom analyzed, especially the mechanical loss arising from the contacting interfaces between the metal mass and the piezoelectric ceramic elements. For this reason, the effect of the mechanical loss and the geometrical dimensions on the mechanical quality factor cannot be studied. In the following analysis, the inner friction loss in the material is neglected and the mechanical loss at the interfaces is represented by a resistance. When the mechanical loss is considered, the effect of the thickness of the thick electrodes and the length and the position of the piezoelectric elements in the sandwich transducer on the transducer performances, such as the resonance and the antiresonance frequency, the mechanical quality factor, and the effective electromechani-

 0001-4966/2004/115(1)/182/5/$20.00

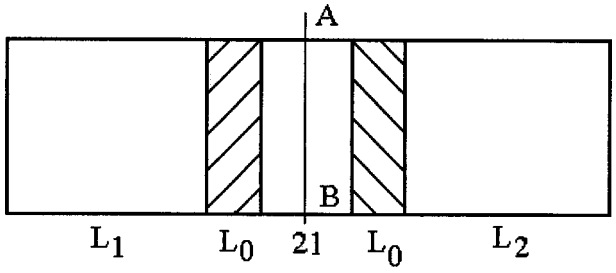


FIG. 1. Schematic of a sandwich piezoelectric ultrasonic transducer with thick electrodes.

cal coupling coefficient are studied, some useful and original results are obtained.

II. DESIGN THEORY FOR THE SANDWICH PIEZOELECTRIC TRANSDUCER WITH THICK ELECTRODES

Figure 1 shows the sandwich piezoelectric transducer with thick electrodes. In Fig. 1, line *AB* represents the displacement node. L_0 is the length of the piezoelectric ceramic elements, L_1 and L_2 are the lengths of the back and the front metal masses and $2l$ is the length of the thick metal electrode. When the transducer is symmetric, we have $L_1=L_2=L$, and the displacement node is located at its geometrical center. For most applications, the sandwich transducer is operated on the fundamental mode. In this case, the transducer may be considered as a combination of two quarter-wavelength vibrators. In the following analysis, only one of the quarter-wavelength vibrators is analyzed. The electromechanical equivalent circuit of the quarter-wavelength vibrator in the sandwich piezoelectric transducer with thick electrodes is shown in Fig. 2. Here, R represents the mechanical loss resistance in the transducer, $Z_{11}=jZ_1\tan(k_1l/2)$, $Z_{12}=Z_1/[j\sin(k_1l)]$, $Z_{01}=jZ_0\tan(k_0L_0/2)$, $Z_{02}=Z_0/[j\sin(k_0L_0)]$, $Z_{21}=jZ_2\tan(k_2L/2)$, $Z_{22}=Z_2/[j\sin(k_2L)]$, $Z_1=\rho_1c_1S_1$, $Z_0=\rho_0c_0S_0$, $Z_2=\rho_2c_2S_2$; $k_1=\omega/c_1$, $k_0=\omega/c_0$, $k_2=\omega/c_2$, $c_1=(E_1/\rho_1)^{1/2}$, $c_2=(E_2/\rho_2)^{1/2}$, $c_0=(1/s_{33}^E\rho_0)^{1/2}$. For a symmetrical transducer, we have, $Z_1=Z_2=Z$, $k_1=k_2=k$; $c_1=c_2=c$. The variables Z, k, c are the characteristic impedance, the wave number, and the sound speed in the front and back metal mass. C_0 is the clamped capacitance of the transducer, and n is the electromechanical conversion coefficient. From Fig. 2, the equivalent mechanical impedance of the quarter-wavelength vibrator is

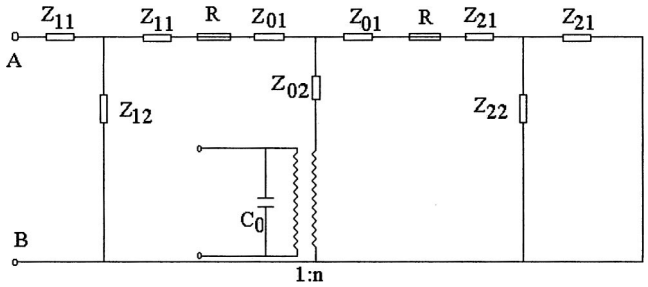


FIG. 2. Electromechanical equivalent circuit of the quarter-wavelength vibrator in the sandwich piezoelectric transducer with thick electrodes.

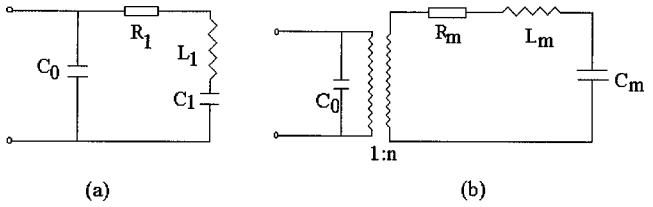


FIG. 3. Lumped parameter electro-mechanical equivalent circuit of the quarter-wavelength vibrator in the sandwich piezoelectric transducer with thick electrodes.

$$Z_m=R_m+jX_m=\frac{AC+BD}{C^2+D^2}+j\left[\frac{BC-AD}{C^2+D^2}-\frac{Z_0}{\sin(k_0L_0)}\right]. \tag{1}$$

In Eq. (1),

$$A=R^2-Z_0^2\tan^2(k_0L_0/2)+ZZ_0\tan(k_0L_0/2)\cot(kl)$$
$$-ZZ_0\tan(kL)\tan(k_0L_0/2)+Z^2\tan(kL)\cot(kl),$$
$$B=2Z_0R\tan(k_0L_0/2)-ZR\cot(kl)+ZR\tan(kL),C$$
$$=2R,\quad D=2Z_0\tan(k_0L_0/2)+Z\tan(kL)-Z\cot(kl).$$

From Fig. 2, the input electrical admittance of the quarter-wavelength vibrator is

$$Y_{ie}=\frac{n^2R_m}{R_m^2+X_m^2}+j\left[\omega C_0-\frac{n^2X_m}{R_m^2+X_m^2}\right]. \tag{2}$$

When the input electrical admittance is infinity, the resonance frequency equation is

$$X_m=0. \tag{3}$$

Setting the imaginary part of the electrical admittance to zero, the antiresonance frequency equation can be obtained as

$$\omega C_0-\frac{n^2X_m}{R_m^2+X_m^2}=0. \tag{4}$$

Solving Eqs. (3) and (4), the resonance and the antiresonance frequencies f_r and f_a can be obtained and the effective electromechanical coupling coefficient k_{eff} can be obtained as the following expression:

$$k_{\text{eff}}^2=\frac{f_a^2-f_r^2}{f_a^2}. \tag{5}$$

At resonance, the lumped parameter electromechanical equivalent circuit of the quarter-wavelength vibrator in the sandwich transducer can be obtained, as shown in Fig. 3. Here, Fig. 3(a) is the electric equivalent circuit; Fig. 3(b) is the electromechanical equivalent circuit. The variables R_m, L_m, C_m are the equivalent mechanical resistance, the equivalent mechanical inductance, and the equivalent mechanical capacitance. In Fig. 3(a). R_1, L_1, C_1 represent the mechanical components reflected onto the electrical terminals. The mechanical quality factor Q_m of the transducer is

$$Q_m=\frac{\omega_rL_m}{R_m}. \tag{6}$$

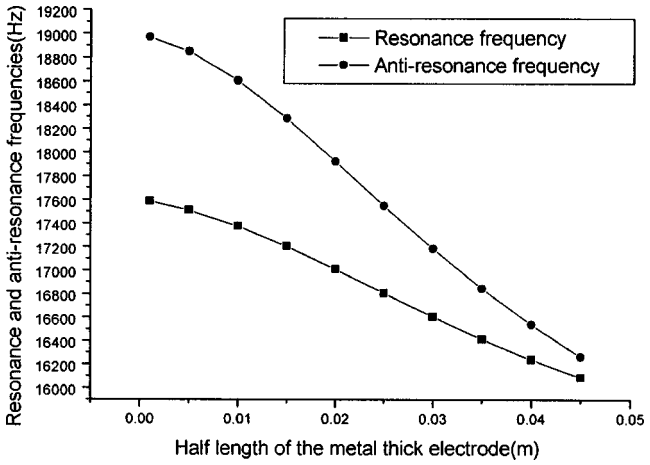


FIG. 4. The resonance frequency and antiresonance frequency versus the thickness of the thick electrodes in the sandwich transducer.

Using the equivalent circuit of the transducer, the resonance angular frequency and the effective electromechanical coupling coefficient can be obtained,

$$k_{\text{eff}}^2=\frac{C_1}{C_0+C_1}, \tag{7}$$

$$\omega_r^2=\frac{1}{L_m C_m}. \tag{8}$$

Using Eqs. (7) and (8), we can get the mechanical quality factor,

$$Q_m=\frac{1-k_{\text{eff}}^2}{k_{\text{eff}}^2}\cdot\frac{n^2}{\omega_r C_0 R_m}. \tag{9}$$

From the above equations, the resonance frequency, the antiresonance frequency, the effective electromechanical coupling coefficient, and the mechanical quality factor can be obtained. However, since the resonance frequency and the antiresonance frequency equations contain transcendental terms, analytical solutions are impossible to find. So a numerical method is used. Figure 4 is the calculated resonance and antiresonance frequencies versus the thickness of the thick electrodes, and Fig. 5 is the calculated effective electromechanical coupling coefficient and the mechanical quality factor versus the thickness of the thick electrodes. In Figs. 4 and 5, the total length of the transducer, i.e., $2L+2l+2L_0$ is kept constant; the length of the piezoelectric ceramic elements is also kept constant.

It can be seen that although the total length of the transducer is kept constant, the resonance frequency of the transducer changes when the lengths of the electrode and the blocks are changed. When the thickness of the electrodes is increased, the resonance frequency, the antiresonance frequency, the effective electromechanical coupling coefficient, and the mechanical quality factor all decrease. Therefore, the thickness of the electrodes must be chosen carefully.

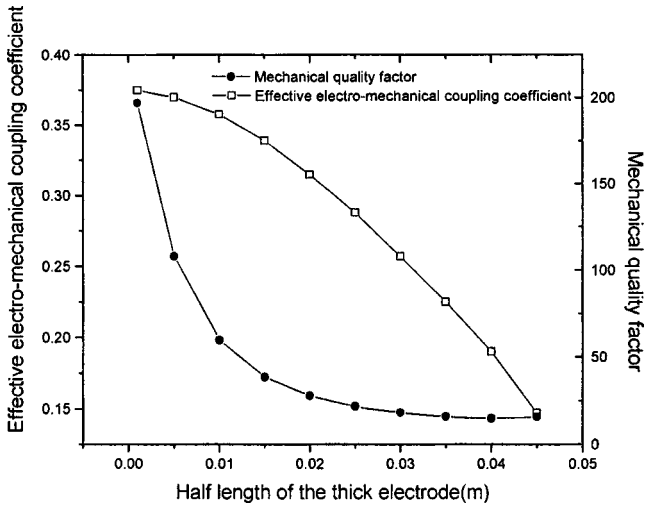


FIG. 5. The effective electromechanical coupling coefficient and the mechanical quality factor versus the thickness of the thick electrodes.

III. EFFECT OF THE LENGTH OF THE PIEZOELECTRIC CERAMIC ELEMENTS ON THE TRANSDUCER PERFORMANCE

It is well known that the piezoelectric ceramic elements affect the transducer performance. In previous studies[6,7] the effect of the piezoelectric elements on transducers was studied. In this section, when the mechanical loss in the transducer is considered, the effect of the length of the piezoelectric elements on the symmetrical sandwich transducer with thin electrodes ($2l=0$) is studied. Using the similar procedure, from Fig. 1 and Fig. 2, the equivalent mechanical impedance of the quarter-wavelength vibrator can be obtained;

$$Z_m=jZ\tan(kL)-jZ_0\cot(k_0L_0)+R. \tag{10}$$

The input electrical admittance of the quarter-wavelength vibrator is

$$Y_{ie}=\frac{n^2R}{R^2+[Z\tan(kL)-Z_0\cot(k_0L_0)]^2}+j\left\{\omega C_0-\frac{n^2[Z\tan(kL)-Z_0\cot(k_0L_0)]}{R^2+[Z\tan(kL)-Z_0\cot(k_0L_0)]^2}\right\}. \tag{11}$$

From Eq. (11), when the input electrical admittance is infinity, the resonance frequency equation can be found,

$$Z\tan(kL)=Z_0\cot(k_0L_0). \tag{12}$$

When the input electrical admittance is zero, the antiresonance frequency equation is

$$\omega C_0=\frac{n^2[Z\tan(kL)-Z_0\cot(k_0L_0)]}{R^2+[Z\tan(kL)-Z_0\cot(k_0L_0)]^2}. \tag{13}$$

Figure 6 and Fig. 7 show the calculated curves of the resonance and antiresonance frequency, the effective electromechanical coupling coefficient, and the mechanical quality factor versus the length of the piezoelectric elements in the symmetrical sandwich transducer with thin electrodes. In the calculations, the total length of the sandwich transducer with thin electrodes is kept constant. When the length of the piezoelectric elements is increased, the length of the back or front metal block is decreased. From Figs. 6 and 7, it can be

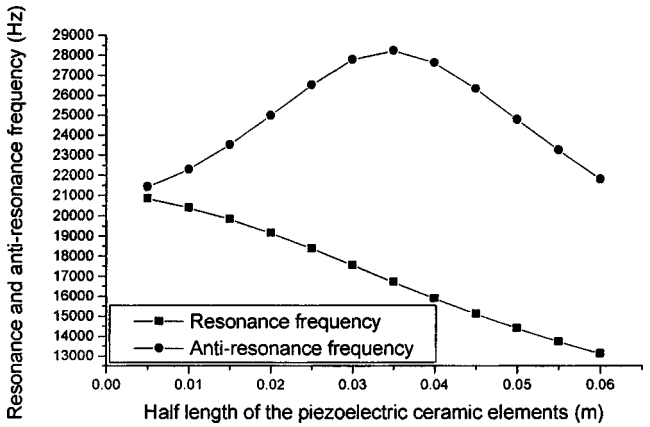


FIG. 6. The resonance frequency and antiresonance frequency versus the length of the piezoelectric elements in the symmetrical sandwich transducer with thin electrodes.

seen that the length of the piezoelectric elements in the transducer affects the transducer performance. When the length of the piezoelectric elements is increased, the resonance frequency and the mechanical quality factor of the transducer are decreased; while the relationships between the antiresonance frequency and the effective electromechanical coupling coefficient with the length of the piezoelectric elements are complex. Corresponding to some definite lengths of the piezoelectric elements, the antiresonance frequency and the effective electromechanical coupling coefficient have maximum values. It should be pointed out that the lengths corresponding to the maximum anti-resonance frequency and the maximum effective electromechanical coupling coefficient are different.

IV. EFFECT OF THE POSITION OF THE PIEZOELECTRIC ELEMENTS ON THE TRANSDUCER PERFORMANCE

For an asymmetric sandwich piezoelectric transducer of thin electrodes, we have $L_1 \neq L_2$, $2l=0$. Since the stress and displacement in the sandwich transducer are different, the position of piezoelectric elements in the transducer will affect the transducer performance. Figure 8 is the equivalent circuit of the half-wavelength asymmetric sandwich transducer with thin electrodes. It should be pointed that Fig. 8 is different from Fig. 2. Figure 2 is the equivalent circuit of the quarter-wavelength transducer; its terminals AB is the mechanical displacement node. In Fig. 8, $Z_{b1}=jZ_1\tan(k_1L_1/2)$, $Z_{b2}=Z_1/[j\sin(k_1L_1)]$, $Z_{01}=jZ_0\tan(k_0L_0/2)$, $Z_{02}=Z_0/[j\sin(k_0L_0)]$, $Z_{f1}=jZ_2\tan(k_2L_2/2)$, $Z_{f2}=Z_2/[j\sin(k_2L_2)]$, $Z_1=\rho_1c_1S_1$, $Z_0=\rho_0c_0S_0$, $Z_2=\rho_2c_2S_2$; $k_1=\omega/c_1$, $k_0=\omega/c_0$, $k_2=\omega/c_2$, $c_1=(E_1/\rho_1)^{1/2}$, $c_2=(E_2/\rho_2)^{1/2}$, $c_0=[1/(s_{33}^E\rho_0)]^{1/2}$, $S_1=\pi R_1^2$, $S_2=\pi R_2^2$, $S_0=\pi(R_0^2-r^2)$, R_1 and R_2 are the radius of the back and front metal masses. R_0 and r are the outer and inner radius of the piezoelectric elements. From Fig. 8, the equivalent mechanical impedance of the half-wavelength asymmetric sandwich transducer with thin electrodes can be obtained; it has the same form as Eq. (1). However, the concrete expressions are different. Here,

$$A=R^2-[Z_0\tan(k_0L_0/2)+Z_1\tan(k_1L_1)]\times[Z_0\tan(k_0L_0/2)+Z_2\tan(k_2L_2)],$$

$$B=R[2Z_0\tan(k_0L_0/2)+Z_1\tan(k_1L_1)+Z_2\tan(k_2L_2)],$$

$$C=2R,$$

$$D=2Z_0\tan(k_0L_0/2)+Z_1\tan(k_1L_1)+Z_2\tan(k_2L_2).$$

Using Eq. (1) and the above equations, the input electric admittance of the half-wavelength asymmetric sandwich

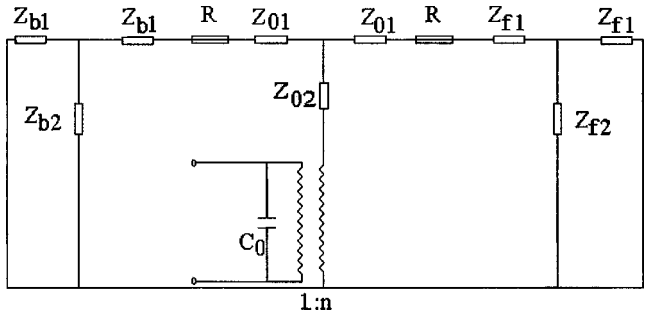


FIG. 8. Equivalent circuit of the half-wavelength asymmetric sandwich piezoelectric transducer with thin electrodes.

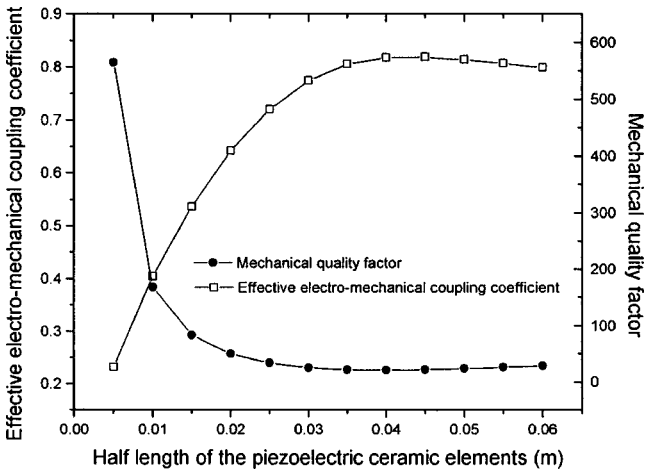


FIG. 7. The effective electromechanical coupling coefficient and the mechanical quality factor versus the length of the piezoelectric elements in the symmetrical transducer with thin electrodes.

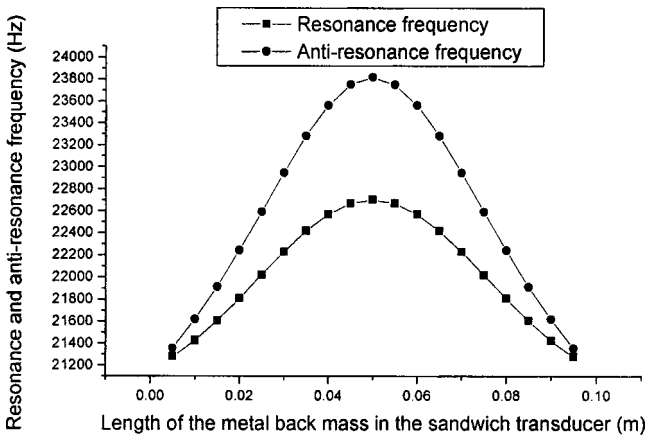


FIG. 9. The resonance frequency and antiresonance frequency versus the position of the piezoelectric elements in the half-wavelength asymmetrical transducer with thin electrodes.

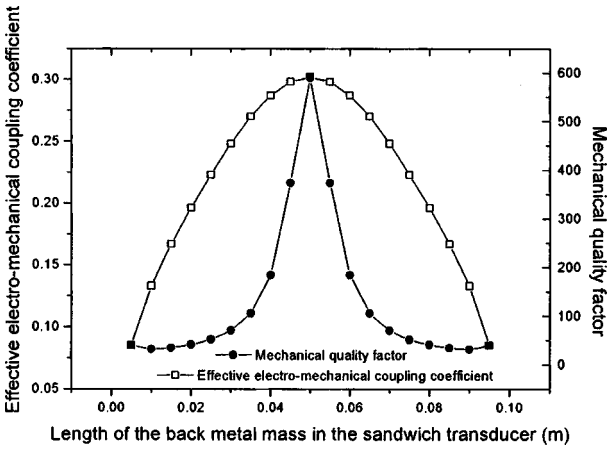


FIG. 10. The effective electromechanical coupling coefficient and the mechanical quality factor versus the position of the piezoelectric elements in the half-wavelength asymmetrical transducer.

transducer can be obtained; the resonance and antiresonance frequency equations can be derived, and the relationship between the position of the piezoelectric elements and the transducer performance can also be calculated numerically. Figure 9 and Fig. 10 show the calculated resonance and antiresonance frequencies, the effective electromechanical coupling coefficient, and the mechanical quality factor versus the position of the piezoelectric elements in the asymmetric sandwich transducer with thin electrodes.

From Fig. 9 and Fig. 10, it can be seen that the position of the piezoelectric elements affects the transducer performance. When the piezoelectric elements are located at the geometrical center, the resonance frequency, the antiresonance frequency, the effective electromechanical coupling coefficient, and the mechanical quality factor have maximum values.

V. CONCLUSIONS

In this paper, the effect of the thickness of thick piezoelectric element electrodes, the length and the position of the piezoelectric elements in the sandwich transducer on the transducer performance is analyzed. To sum up the above analysis, the following conclusions can be obtained.

(1) For high-power sandwich piezoelectric transducers, heating is an important problem. There are many methods that can be used to reduce the heating due to the high excitation voltage and large mechanical strain and stress. One is to use thick electrodes. Since the thickness of the electrodes between the piezoelectric elements is large, the heat can be dissipated easily, and this is beneficial for the high-power ultrasonic transducer. However, the electrodes must not be excessively thick. The reason is that when the thickness of the electrodes is increased, the effective electromechanical coupling coefficient and the mechanical quality factor will decrease. Therefore, in the design of high-power sandwich ultrasonic transducers, the thickness of the electrodes should be chosen carefully.

(2) The length of the piezoelectric elements in the sandwich transducer affects the transducer performance. When the length is increased, the resonance frequency is decreased, while the antiresonance frequency is increased, and there is a maximum value for the antiresonance frequency. On the other hand, when the length is increased, the mechanical quality factor is decreased, while the effective electromechanical coupling coefficient is increased, and there is also a maximum value for the effective electromechanical coupling coefficient. It should be pointed out that the lengths corresponding to the maximum antiresonance frequency and the maximum effective electromechanical coupling coefficient are different.

(3) The position of the piezoelectric elements in the sandwich transducer also affects the transducer performance. When the piezoelectric elements are located at the center of the transducer, the resonance frequency, the antiresonance frequency, the effective electromechanical coupling coefficient, and the mechanical quality factor have maximum values. When the piezoelectric elements deviate from the geometrical center, these parameters will decrease. Therefore, in the design of high-power sandwich ultrasonic transducers, the piezoelectric elements should be placed at the geometrical center for symmetric transducers. For asymmetric transducer, the piezoelectric elements should be placed at the longitudinal displacement node.

ACKNOWLEDGMENT

The author wishes to acknowledge The National Natural Science Foundation of China for its financial support.

[1] M. P. Johnson, "Velocity control and the mechanical impedance of single degree of freedom electromechanical vibrators," J. Acoust. Soc. Am. **84**, 1994–2001 (1988).
[2] R. Coates and R. F. Mathams, "Design of matching networks for acoustic transducers," Ultrasonics **26**, 59–64 (1988).
[3] B. Dubus and J. C. Debus, "Analysis of mechanical limitations of high power piezoelectric transducers using finite element modeling," Ultrasonics **29**, 201–207 (1991).
[4] P. T. Gough and J. S. Knight, "Wide bandwidth, constant beam width acoustic projectors: a simplified design procedure," Ultrasonics **27**, 234–238 (1989).
[5] B. Athikom, K. P. Hari, and D. F. Robert, "Optimizing the performance of piezoelectric drivers that use stepped horns," J. Acoust. Soc. Am. **90**, 1223–1229 (1991).
[6] R. A. Lemaster and K. F. Graff, "Influence of ceramic location on high power transducer performance," *1978 Ultrasonics Symposium Proceedings*, 296–299.
[7] A. Shoh, Sonic transducer, U.S. Patent, No. 3524085, August 1970.

Nonperturbing measurements of spatially distributed underwater acoustic fields using a scanning laser Doppler vibrometer

Andy R. Harland,[a] Jon N. Petzing, and John R. Tyrer
Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, Leicestershire, LE11 3TU, United Kingdom

(Received 22 November 2002; revised 22 October 2003; accepted 31 October 2003)

Localized changes in the density of water induced by the presence of an acoustic field cause perturbations in the localized refractive index. This relationship has given rise to a number of nonperturbing optical metrology techniques for recording measurement parameters from underwater acoustic fields. A method that has been recently developed involves the use of a Laser Doppler Vibrometer (LDV) targeted at a fixed, nonvibrating, plate through an underwater acoustic field. Measurements of the rate of change of optical pathlength along a line section enable the identification of the temporal and frequency characteristics of the acoustic wave front. This approach has been extended through the use of a scanning LDV, which facilitates the measurement of a range of spatially distributed parameters. A mathematical model is presented that relates the distribution of pressure amplitude and phase in a planar wave front with the rate of change of optical pathlength measured by the LDV along a specifically orientated laser line section. Measurements of a 1 MHz acoustic tone burst generated by a focused transducer are described and the results presented. Graphical depictions of the acoustic power and phase distribution recorded by the LDV are shown, together with images representing time history during the acoustic wave propagation. [DOI: 10.1121/1.1635841]

PACS numbers: 43.38.Ne [YHB] Pages: 187–195

I. INTRODUCTION

The propagation of mechanical sound waves in dense media such as water is central to many communication and imaging techniques used in marine and medical applications. The principle of the majority of imaging techniques such as obstetric examination involve relating a measure of the distribution of acoustic energy scattered from boundaries within the volume under interrogation, with a known measure of the acoustic energy provided by the originating source. The quality of the resultant image is largely dependent on the fidelity of these acoustic measurements.

The principal instrument for making underwater acoustic pressure measurements is the hydrophone. A hydrophone typically provides a voltage proportional to the integral of the pressure variation across its active element, although this is generally assumed to be sufficiently small to approximate to a point. Since a hydrophone is required to be submerged in the acoustic field in order to record a measurement, it follows that the field will be perturbed by its presence. Certain design measures, such as ensuring the size of the hydrophone is significantly smaller than the wavelength of the sound being measured and surrounding the hydrophone element with a material of a close impedance matching with water, are used to minimize the perturbation, but it can never be entirely eliminated.

In order to characterize the spatial distribution of an acoustic field using a hydrophone, multiple measurements at positions in three orthogonal dimensions are required, either by repeating single hydrophone measurements or establishing a three-dimensional array of hydrophone elements. Both techniques have shortcomings that restrict their application in many instances, for example, the time required to position and record multiple measurements using a single hydrophone and the likely perturbation of the field caused by a three-dimensional array of hydrophone elements and their mounting structure.

An example of a system that aims to accurately record the spatial distribution of an acoustic pressure field, and is indeed directly traceable to the primary standard, is the UK's National Physical Laboratory (NPL) Ultrasound Beam Calibrator, used in the frequency range 500 kHz to 20 MHz.[1] Here a 21-element membrane hydrophone made from polyvinylidene fluoride is mounted within an acoustic field and pressure measurements are recorded by each element. The position and orientation of the acoustic source is then modified and measurements repeated. The system offers a compromise where field perturbation and the time required to record measurements are both minimized.

The work reported here is aimed at addressing the need for nonperturbing optical techniques capable of recording spatially distributed measurements from underwater acoustic pressure fields. An overview of relevant previously published work follows, which enables the technique described in this work to be placed into context.

A great many researchers have engaged in the prediction, measurement, modeling, and understanding of the inter-

[a]Contact details: A. R. Harland, Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, Leicestershire, LE11 3TU, United Kingdom. Telephone: 01509 227656; fax: 01509 227648; electronic mail: A.R.Harland@lboro.ac.uk

action between light and sound. Based on the initial work of Raman and Nath[2,3] and Debye and Sears[4] on this subject, several techniques for visualizing sound waves have been developed. These include the apparatus developed by Willard for visualizing ultrasonic waves[5] and the application of the schlieren method traditionally used for identifying aberrations within large astronomic objective lenses.[6]

A comprehensive description of the theory of acousto-optic interaction is presented by Klein and Cook.[7] The topic of ultrasonically induced optical diffraction has been the subject of many reviews, two of which were presented by Haran[8] and Monchalin.[9] A more recent summary of high-frequency acoustic measurements by optical techniques was authored by Cook.[10]

The advent of coherent light (laser) sources has enabled a refinement in optical diffraction measurements such that Blomme and Leroy were able to establish the second- and third-order approximation methods for characterizing the diffracted light.[11,12] While each of these techniques offered the ability to provide information regarding the spatial distribution of an acoustic field, the techniques were only able to provide qualitative data.

A departure from the conventional diffraction measurements of light intensity came when Reibold and Kwiek began to measure the amplitude as opposed to the intensity of the light, enabling the complex nature of the various diffracted orders to be investigated. This progression, combined with advances in high-powered computer processing equipment, enabled the development of light diffraction tomography.[13] In this process the magnitude and phase of the acoustic wave were established by filtered back-projection using discrete convolution of the diffracted light at multiple angles through the field. By this method, a single-point measurement of pressure could be derived. Further development of the technique was then documented[14–17] such that in their review of 1995, Reibold and Kwiek described light diffraction tomography as being a capable tool for mapping ultrasonic fields up to 5 MHz in frequency.[18] The technique is, however, limited by the resources and time required to complete a scan and the fact that single erroneous points contribute significantly to the overall uncertainty of the measurement result.

Electronic Speckle Pattern Interferometry (ESPI) was first reported by Leendertz and Butters in 1973[19] and has been applied to the measurement of pressure-induced refractive index changes present in underwater acoustic fields. Having extensively studied refractive index changes induced by airborne acoustic fields emanating from vibrating objects such as loudspeakers using ESPI,[20,21] the Applied Optics Group at the Norwegian University of Science and Technology successfully implemented the observation of underwater acoustic fields using the same technology.[22] In a more specific study, Rustad demonstrated the suitability of ESPI for studying the acoustic field generated by a 3.25 MHz continuous wave medical ultrasound probe, with a spatial resolution of better than 100 μm.[23] This technique is, however, limited by the spatial resolution and fixed aspect ratio of the TV camera used to acquire the images and the fact that the sensitivity of the ESPI system is a complex function of the spherical spreading of the optical wave front and target shape and position.

In a similar approach to that taken by Reibold *et al.* for light diffraction studies, phase stepping and tomographic reconstruction techniques have been applied to measurements of sound fields in air.[24,25] The application of a pulsed laser source has enabled transient acoustic waves to be recorded and tomographically reconstructed from ESPI measurements.[26]

The work described here details a method of using a commercially available laser scanning vibrometer to record noninvasive spatially and temporally distributed measurements from underwater acoustic fields. The work advances previous research by allowing two-dimensional whole field images to be produced from arbitrarily shaped acoustic fields.

II. THEORETICAL ANALYSIS

Laser Doppler vibrometry (LDV) is a well-established optical metrology tool, conventionally used to record surface velocity measurements from vibrating objects.[27] Also known by the more generic term of laser Doppler velocimetry, LDV has been previously demonstrated to be effective in recording the temporal distribution of an underwater pressure wave as it passes over the single line section of the laser beam directed through the acoustic volume.[28–31] Since the instrument used in this work is a commercially available vibrometer, this term is used throughout the work, although it is recognized that the principle by which the instrument records the measurement (by measuring the rate of change of refractive index) is not that by which a conventional vibrometer operates.

Consider the ideal case of a collimated beam of radius r, with plane phase fronts. In its simplest geometry, the beam from the LDV is normal to the axis of the sound beam. In this arrangement, the acoustic phase, Φ, remains constant with distance along the line and the voltage output from the LDV, V, which is proportional to the rate of change of optical path length, is described by Eq. (1), where K is the sensitivity scalar of the LDV electronics, $(\partial n/\partial P)_s$, is the adia-

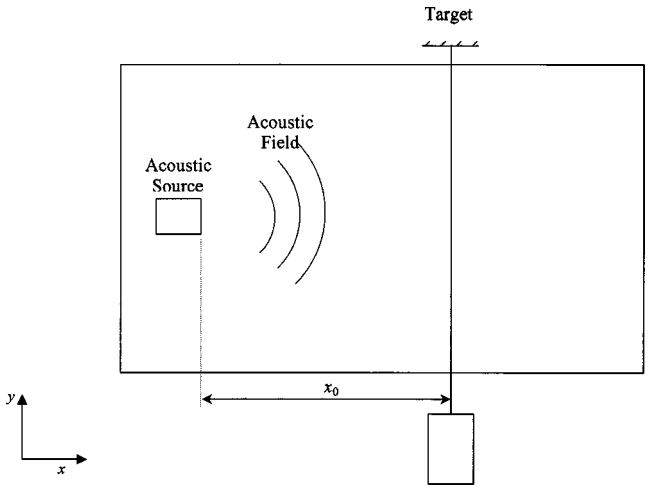


FIG. 1. Experimental arrangement depicting the propagation distance between the acoustic source and the LDV laser beam.

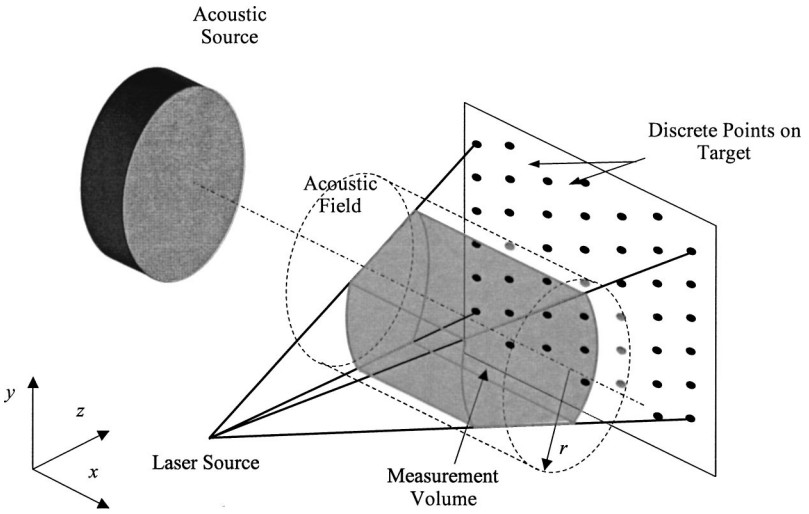


FIG. 2. The interactive range of scanning LDV with the acoustic field, defined by the laser beam at its angular extremities.

batic piezo-optic coefficient, A is the acoustic pressure amplitude, and f is the acoustic frequency. The optical pathlength, l, represents the integral of the refractive index, n, with distance, where the limits of integration are of the path of the laser beam that is affected by the sound field:

$$V(t)=K\frac{dl(t)}{dt}=8\pi rK\left(\frac{\partial n}{\partial P}\right)_s A\cos(2\pi ft-\Phi). \qquad (1)$$

When the LDV measurement is triggered in time by the input voltage to the transducer, the perpendicular distance from the laser beam position to the front face of the acoustic source, x_0, determines the time delay prior to the arrival of the acoustic tone burst. This is depicted in Fig. 1.

It follows, therefore, that information regarding the spatial distribution of the acoustic field can be obtained by comparing the rate of change of the optical pathlength measured at a range of laser beam positions with respect to the acoustic source. An extension to this analysis of a single line section through an acoustic field is to consider multiple line sections simulating the position of a laser beam within a two-dimensional range.

A scanning LDV system allows the beam to be sequentially directed within a range specified by a number of discrete positions established on a fixed, stationary target beyond and outside the acoustic volume, as depicted in Fig. 2. Since each measurement is triggered by the electrical input signal to the source, a referenced measure of spatial and temporal pressure distribution can be established. The output from each measurement recorded in this way is likely to vary in two ways. First in phase, predominantly due to the time-of-flight delay prior to the arrival of the acoustic tone-burst and second in magnitude, determined by the position and incidence angle of the laser beam.

Since the angles of incidence cannot be perpendicular to the acoustic axis for more than one measurement within the same scan, an additional mathematical explanation is needed to accommodate for angular discrepancy. Given that the laser beam is directed by two scanning mirrors, each manipulating the horizontal and vertical position of the beam, respectively, it follows that the position of the beam should be defined by angles in two planes; the x-z plane, represented by ϕ, and the y-z plane represented by θ. A typical arrangement is shown in Fig. 3. The z axis is defined as the line passing from the laser source through the acoustic axis with normal incidence. The y axis remains orthogonal to both the acoustic axis and the z axis.

When the laser beam is incident with arbitrary angles (polar angle ϕ and elevation angle θ, as shown in Fig. 3) on

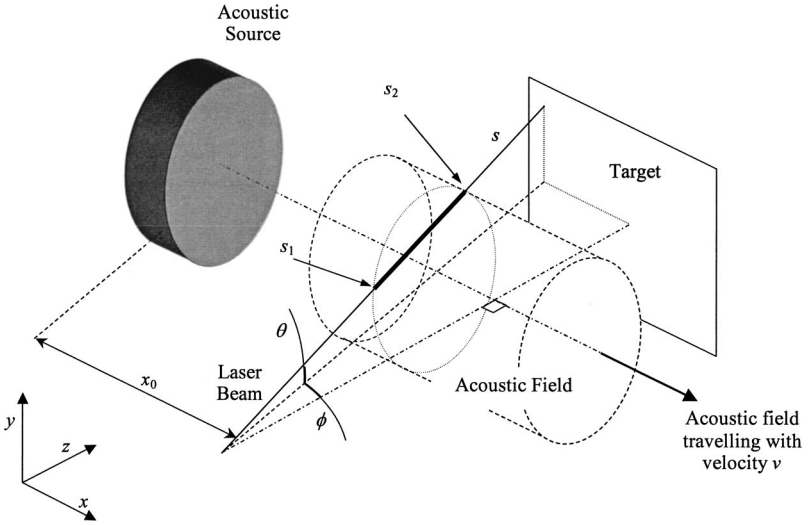


FIG. 3. LDV laser beam passing through an acoustic field with incidence described by the angles ϕ and θ.

the collimated acoustic beam, the optical pathlength, l, is written as

$$l = l_0 + A_0\left(\frac{\partial n}{\partial p}\right)\int_{s_1}^{s_2} \sin[\omega t - k(x_0 + |s|\cos\theta\sin\phi)]ds, \tag{2}$$

where $k=\omega/c$ is the acoustic wave number, and l_0 is the ambient optical pathlength.

The integration leads to

$$l(t) = l_0 + A_0\left(\frac{\partial n}{\partial p}\right)\frac{1}{\alpha}[\cos(\alpha|s_2| - \Psi) - \cos(\alpha|s_1| - \Psi)], \tag{3}$$

where $\alpha = k\cos\theta\sin\phi$ is the wave number projected onto the normal axis, and where $\Psi = \omega t - kx_0$ is the phase term when the beam is normal to the axis of the sound field. The distances $|s_1|$ and $|s_2|$ are indicated in Fig. 3. They represent the position at the beginning and the end of the acousto-optic interaction region, respectively, and are defined in Cartesian vector form in Eq. (4), where d_a is the distance of the acoustic source from the laser source:

$$\begin{aligned}\overline{s_1} &= \begin{bmatrix} s_{1x} \\ s_{1y} \\ s_{1z} \end{bmatrix} \\ &= \begin{bmatrix} \tan\phi\cos\theta(d_a\cos\theta - \sqrt{r^2 - (d_a\sin\theta)^2}) \\ \sin\theta(d_a\cos\theta - \sqrt{r^2 - (d_a\sin\theta)^2}) \\ \cos\theta(d_a\cos\theta - \sqrt{r^2 - (d_a\sin\theta)^2}) \end{bmatrix},\end{aligned} \tag{4}$$

$$\begin{aligned}\overline{s_2} &= \begin{bmatrix} s_{2x} \\ s_{2y} \\ s_{2z} \end{bmatrix} \\ &= \begin{bmatrix} \tan\phi\cos\theta(d_a\cos\theta + \sqrt{r^2 - (d_a\sin\theta)^2}) \\ \sin\theta(d_a\cos\theta + \sqrt{r^2 - (d_a\sin\theta)^2}) \\ \cos\theta(d_a\cos\theta + \sqrt{r^2 - (d_a\sin\theta)^2}) \end{bmatrix},\end{aligned}$$

such that the respective distances from the laser source to the points s_1 and s_2 are given by

$$\begin{aligned}|s_1| &= \sqrt{-\frac{(-d_a\cos\theta + \sqrt{r^2 - d_a^2 + d_a^2\cos^2\theta})^2(-\cos^2\theta + \cos^2\theta\cos^2\phi - \cos^2\phi)}{\cos^2\phi}}, \\ |s_2| &= \sqrt{-\frac{(d_a\cos\theta + \sqrt{r^2 - d_a^2 + d_a^2\cos^2\theta})^2(-\cos^2\theta + \cos^2\theta\cos^2\phi - \cos^2\phi)}{\cos^2\phi}}.\end{aligned} \tag{5}$$

Differentiating the total optical pathlength, $2l$, with respect to time to calculate the rate of change of optical pathlength as measured by the LDV, gives the following:

$$\begin{aligned}\frac{dl(t)}{dt} = 2A_0\left(\frac{\partial n}{\partial p}\right)\frac{\omega}{\alpha}[&\sin(\omega t - (kx_0 - |s_1|\alpha)) \\ &- \sin(\omega t - (kx_0 - |s_2|\alpha))].\end{aligned} \tag{6}$$

From Eq. (6) it can be seen that the sensitivity of a LDV laser beam is dependent on its angular position with relation to the acoustic axes, θ and ϕ, and the frequency, f, of the acoustic wave. Due to the presence and relative positions of these variables within Eq. (6), it follows that at certain angular positions, the pressure distribution will be such that the integral of refractive index with distance along the laser beam through the interactive region is equal to zero. This will result in the optical pathlength of the laser beam remaining constant with time. Hence, at certain pairs of angles θ and ϕ, the LDV, which provides a measure of the rate of change of optical pathlength, will be totally insensitive to the acoustic wave.

One measure that can be taken to minimize the influence of the angular position of the beam is to ensure that the stand-off distance between the scanning head and the acoustic field is sufficiently greater than the range of the scan to enable each beam position to be approximately parallel.

The theoretical rate of change of optical pathlength described in Eq. (6) is valid only for LDV interrogations of continuous wave acoustic signals. These are not generally used in laboratory-based underwater acoustic studies due to the need to eliminate interference effects between the signal and other portions of the signal reflected from the tank walls, surfaces, or other impeding obstacles. For this reason, tone-burst acoustic signals are used. These consist of typically between 5 and 50 cycles of a single frequency separated by a delay sufficient for the reflected components of the wave to die away.

While the mathematical description of the rate of change of the optical pathlength given in Eq. (6) holds true for the scenario shown in part (b) of Fig. 4, where the interactive region of the laser beam extends from the near to the far side of the acoustic field, it does not describe the measured effect depicted in parts (a) and (c). In both these scenarios, the interactive region is reduced from that of the mathematical description.

To account for the period before and after interaction with the steady-state region of the acoustic wave, one of two approaches can be taken. The first is to consider the tone burst to be a continuous wave but relate the limits of the integral, s_1 and s_2, to the position of the tone burst in time and space, such that the distance over which the integral is taken, $(s_2 - s_1)$, is equal to zero before and after the acoustic

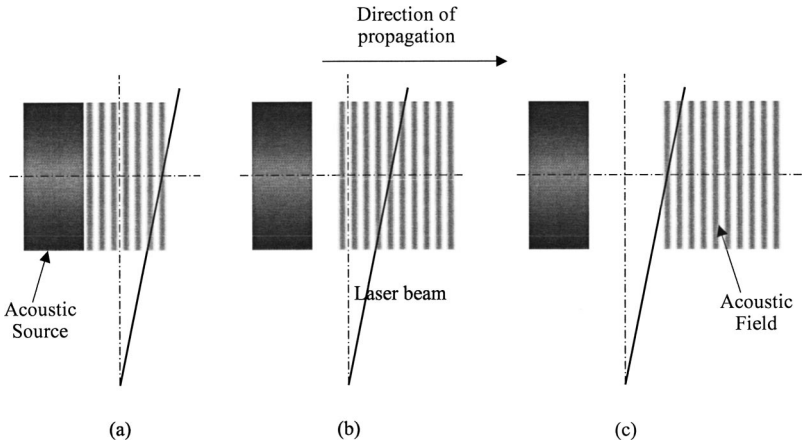


FIG. 4. LDV laser beam traversing a propagating acoustic tone burst.

tone burst has passed. The second approach is to modify the description of pressure from a continuous wave into a tone burst by multiplying the signal by a rectangular window of an appropriate duration in the time domain.

In practice, it is extremely unlikely that the wave front from a plane–piston acoustic source will exhibit entirely plane wave characteristics. Instead, the amplitude, phase, shape, and divergence of the acoustic field will mean that the description of pressure variation with time at a point will be significantly more complex than that used for the mathematical description presented so far. In addition, there are many other designs of acoustic transducer in industrial use, each of which enables the generated field to exhibit certain characteristics.

One such design in widespread use in the medical community is the focused transducer. Here, a concave transducer face generates a spherical wave front, which reduces to a minimum cross section at a focal point some distance away from the transducer. The relationship between the amplitude and phase of the pressure variation with spatial position in three dimensions in a focused field would require the mathematical description of the rate of change of an optical pathlength given in Eq. (6) to take a more generic form, as given in Eq. (7):

$$\frac{dl(t)}{dt} = 2\frac{d}{dt}\left[\left(\frac{\partial n}{\partial p}\right)\int_{s_1}^{s_2} A(s)\cos(\omega t - \Phi(s))ds\right], \qquad (7)$$

where $A(s)$ is the amplitude and $\Phi(s)$ is the phase, both as an unknown function of distance, s.

Given the presence of unknown amplitude and phase components within Eq. (7), it follows that without knowledge of the acoustic field, it is impossible to derive pressure amplitude and phase components from experimentally measured data. It is also recognized that in the experimental results that follow, no field can be considered perfectly planar. However, this does not appear to hinder the measurement technique, which provides acceptable visualizations from nonuniform fields.

This mathematical analysis is presented to assist in the interpretation of the results that follow. Since each measurement represents an integral of the refractive index with distance, care should be taken when considering quantities such as acoustic pressure with respect to the experimental results.

III. EXPERIMENTAL RESULTS

The experimentation reported here was carried out using a PSV-300 series scanning LDV system manufactured by *Polytec* GmbH. The system consisted of a scanning head from which the optical measurement laser beam was emitted in a direction controlled by two orthogonal scanning mirrors, associated hardware, and PC-based control software. Integrated within the scanning head was a Charge Couple Device (CCD) TV camera, which facilitated remote setup and provided reference images onto which graphical results could be superimposed.

The experimental arrangement was set up as shown in Fig. 5, with the LDV scanning head positioned approximately 1 m from the acoustic axis. The laser beam was passed through the front glass wall of the tank, through the measurement volume, through the rear glass wall, and returned along the same path from the stationary target. The target consisted of a rigid plate upon which was mounted a 100 mm×100 mm panel of retroreflective material, consisting of a layer of microscopically smooth glass beads adhered to a mirrored backing. A grid of specified increments in x and y was then established on the target, the nodes of which defined the measurement positions for the laser beam. The angular deflections of each of the scanning mirrors were automatically recorded at each of the nodal measurement position.

At each target position, a measurement of the rate of

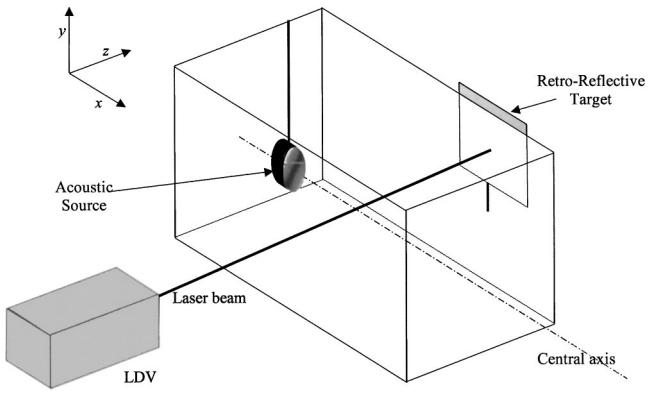


FIG. 5. Experimental arrangement used to record measurements from underwater acoustic fields.

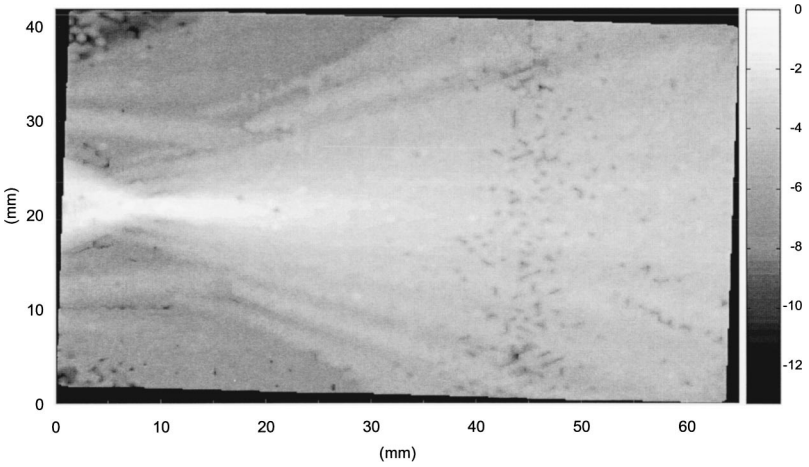


FIG. 6. ''Power'' distribution of the rate of change of the optical pathlength ($\log m^2/s^2$).

change of the optical pathlength was recorded, triggered in time from the electrical input signal to the acoustic source. The distance of the measurement position from the source, the acoustic frequency, and the number of acoustic cycles determined the measurement duration. The signal was sampled at 40 MHz and a Fast Fourier Transform (FFT) with a maximum resolution of 6400 lines was recorded in software.

Since the conventional application of a LDV is to record velocity measurements from a target, the software provided with the scanning system is programed to interpret the measured data as target velocity measurements, which was not the case in the experiments described here. These data are then used to provide a range of graphical representations of the motion and frequency response of the target. For this reason, the original rate of change of pathlength data recorded by the vibrometer were extracted and processed using *Matlab* to enable measurements of the acoustic field within the water to be derived.

The color or gray-scaled images are widely used to represent the spatial distribution of any distributed parameter, where a specific value representative of the quantity under scrutiny is calculated for each measurement position and presented as a shaded pixel. When each pixel is positioned relative to the neighboring measurements, the array of individual elements combines to form the image. A graduated scale is then calibrated from the lowest to the highest pixel value contained within the image.

Data recorded from the LDV were used to generate images representative of a number of measured quantities. The previously recorded angular positions were used to position each measurement point within the final image. A linear interpolation was then undertaken between adjacent measurements to increase the number of pixels in each axis by a factor of 5, thus improving the visual quality of the images.

In order to provide an example of the range of information that can be extracted from each LDV scan, an experimental scan was completed of the field generated by a Panametrics V3438 transducer emitting a 1 MHz sinusoidal tone burst consisting of five cycles.

A scanning grid was defined containing a total of 7543 discrete laser target points was established. The grid consisted of 71 odd columns each of 53 points and 70 even columns each of 54 points, offset vertically from the odd columns by half the point separation in an ''isometric'' arrangement, similar to that depicted in Fig. 2. An average of 15 successively triggered time-swept signals of 50 μs duration were recorded at a sampling frequency of 5.12 MHz, and 400 FFT lines stored in a Universal File format.

A measure of the magnitude or power at a certain frequency within a signal measured using a LDV in this way was established from the respective FFT component of the magnitude or power spectrum at the frequency concerned. Each complex FFT recorded by the LDV was converted into a power spectrum and the component at the fundamental acoustic frequency was taken to represent the ''power'' of the signal. The ''magnitude'' was calculated as the square root of the calculated power value. Both magnitude and power are quantities derived from the rate of change of the optical pathlength or velocity and take the units of m/s. Given only the data recorded by the LDV, it is not possible to derive parameters such as acoustic pressure, intensity, or particle displacement, due to the integration effect previously described. It is anticipated that further research and investigation will provide methods for overcoming this shortcoming. Therefore, to avoid a misunderstanding due to idiosyncratic units, it is convenient to normalize the data of these images, creating dimensionless values between 0 and 1. Images were presented on a logarithmic scale, since this was found to depict their quantity with improved visual clarity.

The principal reason for the use of tone-burst signals in laboratory-based underwater acoustics is to enable measurements to be made without interference from acoustic components reflected from the walls or floor of the water tank or upper air–water interface. Therefore, for each measurement, the ''power'' and ''magnitude'' values were extracted from a 4 μs portion of the signal coinciding with the center of the acoustic tone burst to minimize the influence of spurious acoustic components. The resultant ''power'' image is given in Fig. 6.

In this particular example, the reflecting target contained a surface defect (scratch) running from the top to the bottom of the scan approximately one-quarter of the image width from the right hand side. This is the cause of the noisy band observed in this region.

In addition to the ''power'' or ''magnitude'' within the

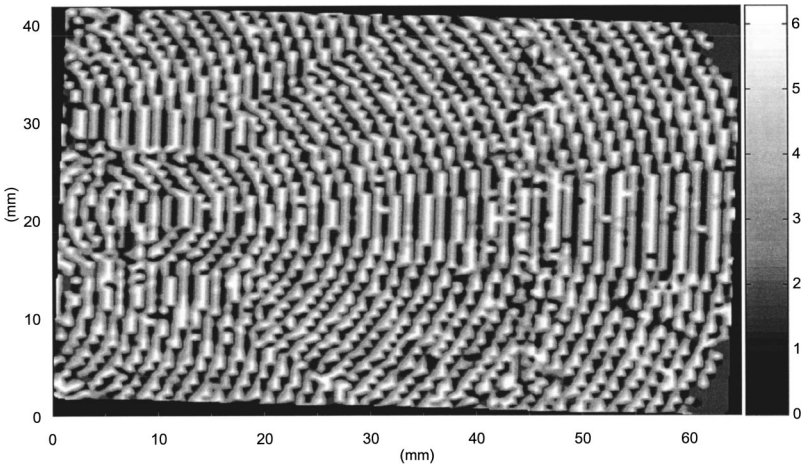


FIG. 7. ''Phase'' distribution of the rate of change of the optical pathlength (rads).

field, it is important to consider the phase of the amplitude in order to gain a complete understanding of the nature of the acoustic field distribution. Phase data can also be extracted from the measurements; either in terms of the absolute phase—calculated from the arctangent of the quotient of the imaginary component divided by the real component of the FFT—or as a relative phase by converting the position of the tone-burst signal in time with respect to an adjacent point into phase using the period of the signal. The second of these approaches was found to yield the highest quality ''phase'' distribution images, as shown in Fig. 7.

The most significant observation from the ''phase'' distribution image is the effect of the spatial sampling resolution on the continuity of the phase distribution. As an absolute minimum, Nyquist's sampling theory states that at least two sampling points per wavelength are required, although it is desirable for this value to be nearer 10. Given that the LDV interrogation grid contained 141 discrete columns, the 42 complete cycles observed in the measurement suggest that the spatial sampling rate is toward the lower end of the acceptable range. This is evident in the number of apparent discontinuities in certain regions of the image.

From the derived ''magnitude'' and ''phase'' quantities, an image can be created depicting a continuous wave acoustic wave passing through the field at a nominal time by multiplying the ''magnitude'' by the sine of the phase. Images of this type are able to illustrate the relationship between the magnitude and phase throughout the scan region. By creating a sequence of these images each with a phase shift from the previous image, a cyclic animation mimicking the motion of a continuous acoustic wave through the field can be created.

Using the time-resolved rate of change of the optical pathlength data at each point, a three-dimensional array representing the rate of change of the optical pathlength in *x*, *y*,

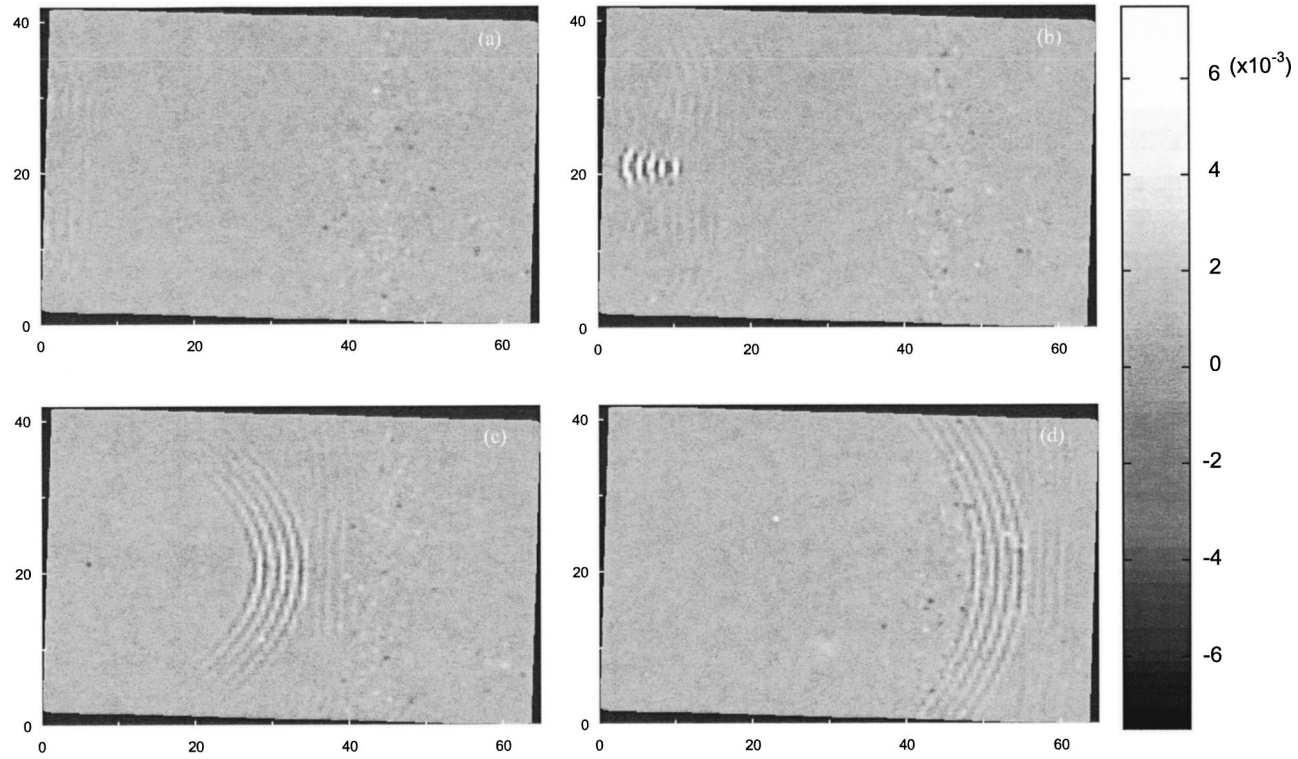


FIG. 8. Time-resolved rate of change of the optical path length (m/s).

and time, t, can be established. By extracting a plane of data in x and y at a given time, t_n, a depiction of the position and amplitude of the acoustic tone burst can be obtained at that instant. Four images depicting the acoustic tone burst at different time instants are given in Fig. 8.

The four instants in time at which the acoustic tone burst is depicted in Fig. 8 are 1.5, 50, 130, and 200 μs after the beginning of the input signal to the acoustic source in (a), (b), (c), and (d), respectively.

Given the ability of the post-processed scanning LDV measurements to enable reconstructions of images at instants in time, it follows that a series of sequential images can be created, which form the individual frames of an animated picture. Consequently, for each LDV scan, an animation can be produced, recreating the passage of an acoustic tone burst through the scanning region. Without interpolation in the time domain, the maximum number of frames constituting the animation is equal to the number of time samples of the recorded signal. An animation created in this way differs from one that can be obtained from the magnitude and phase distribution, in that it represents the true acoustic tone burst, as opposed to a simulated continuous wave.

Previous work has characterized the performance of the *Polytec* scanning LDV in terms of its minimum resolvable signal amplitude, calculated to be 18.9 mPa/$\sqrt{\text{Hz}}$,[29] and its agreement with the NPL Laser Interferometer, measured to be within 2.5% at 1 MHz.[29] The same LDV has been used throughout the experiments reported in this work.

IV. CONCLUSIONS

The work reported here describes a nonperturbing optical method for recording spatially distributed measurements of an underwater acoustic field. The instrument has been demonstrated to record optical pathlength changes with time along a finite line section of very small diameter ($<$1 mm). By utilizing the refractive index changes induced by the acoustic field, the presence and spatial and temporal distribution of the acoustic tone burst can be clearly identified and quantitatively measured. It has been previously demonstrated that the minimum resolvable signal amplitude is 18.9 mPa/$\sqrt{\text{Hz}}$.[29]

A mathematical model has been developed to enable the passage of a LDV beam through an acoustic field to be analyzed. The model is built on well-established physical theory of acousto-optic interaction and developed to cater for the case of a heterodyne or velocity interferometer such as a LDV. The theory has been advanced to permit interrogation at any arbitrary angle of incidence between the optical and acoustic beams

A series of mathematical manipulation and extraction enable images representing the acoustic ''power'' and ''phase'' within the acoustic field to be generated. These images provide excellent spatial detail of features, patterns, and discontinuities within the field. Additionally, images depicting the position and magnitude of an acoustic tone burst at an instant in time can be generated.

It should be noted that while a number of quantitative measures of acoustic parameters can be recorded from a repeatable, arbitrary shaped field, at this stage it is not possible to derive acoustic pressure or intensity at a specific point.

The application of a scanning LDV system has been demonstrated to offer wide-ranging possibilities for the measurement of the spatial characterization of an acoustic field. The ability of the scanning LDV system to rapidly record data from a series of discrete line sections through the field was a significant factor in the establishment of this method as a valuable tool for use in underwater acoustic measurement. It has been successfully demonstrated that the technique allows spatially resolved measurements to be made, which illustrate field components that might otherwise have not been observed.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of Roy Preston and Stephen Robinson from the National Physical Laboratory, UK, for their assistance with the experimentation reported in this work and to Roger Traynor of Lambda Photometrics for the loan of the Polytec scanning LDV and similar equipment.

[1] R. C. Preston, ''The NPL ultrasound beam calibrator,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 122–139 (1988).
[2] C. V. Raman and N. S. N. Nath, ''The diffraction of light by high frequency sound waves: Part I,'' Proc. Indian Acad. Sci. **2**, 406–412 (1935).
[3] C. V. Raman and N. S. N. Nath, ''The diffraction of light by high frequency sound waves: Part II,'' Proc. Indian Acad. Sci. **2**, 413–420 (1935).
[4] P. Debye and F. Sears, ''On the scattering of light by supersonic waves,'' Proc. Natl. Acad. Sci. U.S.A. **18**, 409 (1932).
[5] G. W. Willard, ''Ultra-sound waves made visible,'' Bell Lab. Rec. **5**, 194–200 (1947).
[6] A. Toepler, ''Beobachtungen nach einer neuen optischen Methode,'' Max Cohen und Sohn, Bonn, 1864.
[7] W. R. Klein and B. D. Cook, ''Unified approach to ultrasonic light diffraction,'' IEEE Trans. Sonics Ultrason. **14**, 123–134 (1967).
[8] M. E. Haran, ''Visualization and measurement of ultrasonic wavefronts,'' Proc. IEEE **67**, 454–466 (1979).
[9] J.-P. Monchalin, ''Optical detection of ultrasound,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **33**, 485–499 (1986).
[10] B. D. Cook, ''The measurement of high frequency sound fields by optical techniques,'' SPIE **2358**, 278–280 (1994).
[11] E. Blomme and O. Leroy, ''Diffraction of light by ultrasound: Closed analytical expression for the spectra up to order 2,'' J. Acoust. Soc. India **11**, 1–6 (1983).
[12] E. Blomme and O. Leroy, ''Diffraction of light by ultrasound: finite analytical expressions for the spectra up to order 3,'' Acoustica **57**, 168–174 (1985).
[13] R. Reibold and W. Molkenstruck, ''Light diffraction tomography applied to the investigation of ultrasonic fields. Part I: Continuous waves,'' Acustica **56**, 180–192 (1984).
[14] R. Reibold and P. Kwiek, ''On ultrasound light diffraction,'' Ultrasonics **31**, 307–313 (1993).
[15] R. Reibold and P. Kwiek, ''Extension of light diffraction tomography beyond the weak acousto-optic interaction,'' Acustica **81**, 43–52 (1995).
[16] P. Kwiek and R. Reibold, ''Determination of an ultrasound field using light diffraction,'' Acustica-acta acustica **82**, 376–378 (1996).
[17] R. Reibold and P. Kwiek, ''Uncertainty considerations of ultrasonic field mapping by light diffraction tomography,'' Ultrasonics **35**, 187–193 (1997).
[18] R. Reibold and P. Kwiek, ''Ultrasound field mapping by light-diffraction tompgraphy: A review,'' SPIE **2643**, 66–79 (1995).
[19] J. A. Leendertz and J. N. Butters, ''An image shearing speckle pattern interferometer for measuring bending moments,'' J. Phys. E **6**, 1107–1110 (1973).
[20] O. J. Lokberg, ''Sound in flight: measurement of sound fields by use of TV holography,'' Appl. Opt. **33**, 2574–2584 (1994).
[21] O. J. Lokberg, ''Recording of sound emission and propagation in air using

TV holography,'' J. Acoust. Soc. Am. **96**, 2244–2250 (1994).
[22] R. Rustad, O. J. Lokberg, H. M. Pedersen, K. Klepsvik, and T. Storen, ''TV holography measurements of underwater acoustic fields,'' J. Acoust. Soc. Am. **102**, 1904–1906 (1997).
[23] R. Rustad, ''Acoustic field of a medical ultrasound probe operated in continuous wave mode investigated by TV holography,'' Appl. Opt. **37**, 7368–7377 (1998).
[24] O. J. Lokberg, M. Espeland, and M. Pedersen, ''Tomography reconstruction of sound fields using TV holography,'' Appl. Opt. **34**, 1640–1645 (1995).
[25] M. Espeland, O. J. Lokberg, and R. Rustad, ''Full field tomographic reconstruction of sound fields using TV holography,'' J. Acoust. Soc. Am. **98**, 280–287 (1995).
[26] P. Gren, S. Schedin, and X. Li, ''Tomographic reconstruction of transient acoustic fields recorded by pulsed TV holography,'' Appl. Opt. **37**, 834–840 (1998).
[27] D. C. Williams, in *Optical Methods in Engineering Metrology* (Chapman and Hall, London, 1993), ISBN 0-412-39640-8.
[28] A. R. Harland, J. N. Petzing, and J. R. Tyrer, ''Non-invasive measurements of underwater pressure fields using laser Doppler velocimetry,'' J. Sound Vib. **252**, 169–177 (2002).
[29] A. R. Harland, J. N. Petzing, J. R. Tyrer, C. J. Bickley, S. P. Robinson, and R. C. Preston, ''Application and assessment of laser Doppler velocimetry for underwater acoustic measurements,'' J. Sound Vib. (in press).
[30] A. R. Harland, J. N. Petzing, and J. R. Tyrer, ''Assessment and Application of LDV to underwater acoustic measurements,'' *Proceedings of Photon '02—Applied Optics and Optoelectronics Conference*, Cardiff, UK, September 2002, pp. 47–48.
[31] A. R. Harland, J. N. Petzing, and J. R. Tyrer, ''Non-invasive measurements of spatially and temporally distributed underwater acoustic pressure fields using laser Doppler velocimetry,'' accepted for presentation, Institute of Acoustics Conference, Teddington, January 2003.

Surface fields on nonseparable geometries

Angie Sarkissian[a)]
Naval Research Laboratory, Washington, DC 20375-5350

(Received 23 May 2003; accepted for publication 20 October 2003)

The field on the surface of a structure having nonseparable geometry may be expanded in a set of eigenfunctions suitable for that particular geometry. The eigenfunctions are computed numerically using the boundary elements method to compute a matrix that relates the field on the surface of the structure to its normal derivative at vanishing frequency and diagonalizing the matrix. The field on the surface of a finite cylinder with hemisphical end-caps is expanded in such a set of eigenfunctions for both a simple cylinder and a cylinder containing frames. Results for the framed cylinder is compared to Fourier decomposition of the field on the cylindrical portion of the structure.
[DOI: 10.1121/1.1632483]

PACS numbers: 43.40.Fz, 43.20.Fn, 43.20.Ks [EGW] Pages: 197–199

I. INTRODUCTION

When studying the acoustic behavior of a structure having a separable geometry, it is convenient to expand the surface field in the normal modes of the structure to gain understanding. When studying the field produced by a long, finite cylinder, the expansion of the surface field on the cylindrical portion of the structure alone in Fourier modes sheds light on its behavior.[1–3] In order to include the contribution of the end-caps or to examine the behavior of an arbitrarily shaped structure, we describe a set of eigenfunctions to use as an expansion for the surface field. Eigenfunctions suitable for the particular geometry may be computed numerically using the boundary elements method to compute a matrix that relates the acoustic field on the structure surface to the normal derivative of the field at vanishing frequency and diagonalizing that matrix. The decomposition of the surface field in this set of eigenfunctions is applied to a simple hemispherically end-capped cylindrical shell and a framed cylindrical shell. Results are compared to Fourier mode decomposition of the field on the cylindrical region of the surface alone.

II. EIGENFUNCTIONS

The functions used here were originally developed for acoustic holography.[4] They are generated numerically by discretizing the structure surface and applying the boundary elements method to generate a matrix that relates the field on the surface of the structure to its normal derivative for a field satisfying the Laplace equation, which is the acoustic field at vanishing frequency. The matrix so obtained is real and symmetric. The eigenfunctions of this matrix provide a suitable expansion for the field on the structure surface. Ideally it would be more appropriate to compute an impedance matrix at the frequency under consideration and diagonalizing it. But when working with broadband data, the computation of a separate set of functions at each frequency is computationally intensive. The eigenfunctions computed from the Laplace equation provide a suitable set at all frequencies for high mode numbers, as discussed in Refs. 4 and 5. The disadvantage of this expansion is that very low mode numbers become less meaningful.

We consider the scattering problem at end-incidence where both the structure and the acoustic field are symmetric about the z axis, thus circumferential mode number 0 is only present. The details of computation of the eigenfunctions are given in Refs. 4 and 5 and will not be repeated here. Figure 1 shows modes 21–25 for circumferential mode 0 for a cylinder having a total length to a diameter ratio of 7. The eigenfunctions are plotted as a function of normalized arclength. The vertical dashed lines separate the end-caps from the cylindrical region of the structure. The eigenfunctions resemble sines and cosines in the cylindrical portion of the structure, in the middle, and Legendre polynomials in the hemispherical end-cap regions. The functions are ordered such that as the mode number increases, the number of nodes increases so that the mode number is approximately proportional to the wavenumber on the cylindrical portion of the structure.

III. SURFACE FIELD

First the scattered field is computed for a simple hemispherically end-capped cylinderical shell having a total length (including end-caps) to diameter ratio of 7. We give the dimensions of the cylinder in terms of its radius a. The thickness is $0.0074a$. The parameters of nickel are used for the shell, where the Young's modulus is 2.1×10^{11} Pa, the Poisson's ratio is 0.3, the density is 8800 kg/m^3 and the loss factor is 0.001. The sound speed in the fluid exterior to the cylinder is 1510 m/s.

The surface field is computed using the SARA-2D[6] program which uses finite elements to model the structure and both finite and infinite elements to model the fluid medium outside the structure.

Figure 2 shows eigenfunction decomposition of the surface normal velocity for $3<ka<50$, where k is the wavenumber in the exterior fluid. The amplitudes are normalized to reach a maximum of unity. The color scale reaches a maximum of 0.09. Amplitudes higher than 0.09 appear red. The first 16 modes are not shown because they are not mean-

a)Electronic mail: angie@pa.nrl.navy.mil

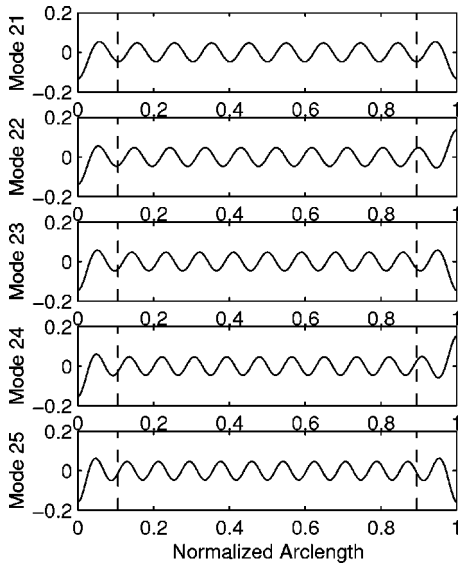


FIG. 1. Eigenfunctions 21–25 are shown as a function of normalized arclength. The dashed lines separate the end-caps from the cylindrical regions.

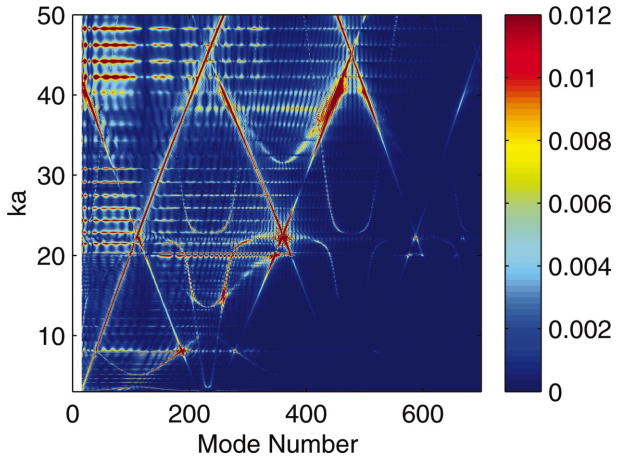


FIG. 3. Eigenfunction amplitudes for a framed hemispherically end-capped cylinder.

ingful in this representation. For this simple shell we observe three waves. The strongest is the wave in the middle traveling at the speed of the acoustic wave. Since the mode number is approximately proportional to wavenumber, this wave appears approximately as a straight line. The slower flexural wave to the right appears approximately as a parabola since its speed is proportional to the square root of frequency and, to the left, the faster compressional wave can be observed as a faint line. The transverse membrane wave does not appear here because of the axial symmetry of the structure and the incident field.

The decomposition is next applied to a framed shell having the same dimensions and material parameters but, in this case, containing 85 identical straight frames along its cylindrical portion, with equal spacing between them. The frames have a length of $0.078a$ and a thickness of $0.0065a$.

Figure 3 shows the eigenfunction decomposition of the surface normal velocity for the same frequency range, $3<ka<50$. We compare it to Fig. 4, showing a Fourier decomposition on the cylindrical portion of the structure. The amplitudes here are also normalized to reach a maximum of unity. The color scale here reaches a maximum of 0.012 with all amplitudes higher than 0.012 appearing as red. The Fourier decomposition contains positive and negative mode numbers describing traveling waves moving to the right or to the left.

The strong wave traveling at the speed of the acoustic wave appears in both figures as a straight line. In Fig. 4 we see lines parallel to it produced by the periodicity of the ribs. Since Fig. 3 represents eigenfunction decomposition rather than traveling waves, we observe in it lines that are similar to the ones that appear on the right side of Fig. 4, having positive Fourier mode numbers, as well as the mirror image of lines that appear on the left side of Fig. 4, having negative mode numbers.

The parabolic flexural wave appearing in Fig. 2 is curved and split into passbands and stopbands in Figs. 3 and 4. The compressional wave that appears as a straight line in Fig. 2 is curved in Figs. 3 and 4 and appears with periodicity in the horizontal direction as does the flexural wave. This

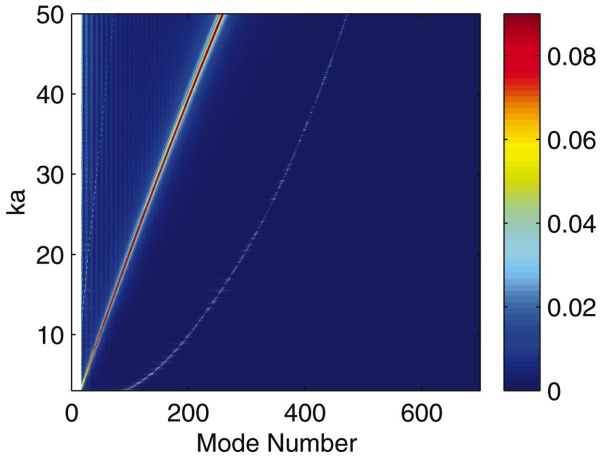


FIG. 2. Eigenfunction amplitudes for a simple hemispherically end-capped cylinder.

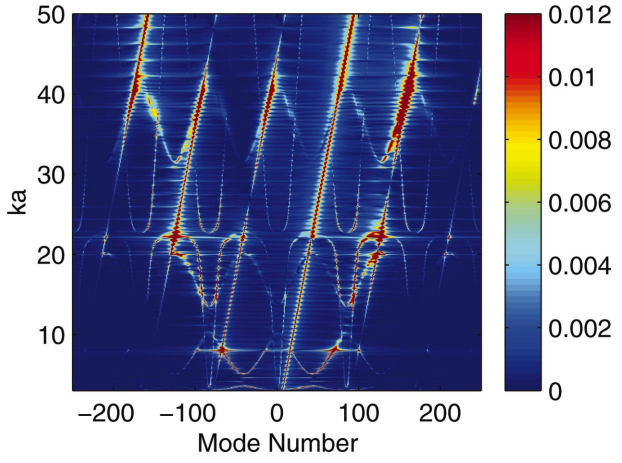


FIG. 4. Fourier mode decomposition for a framed hemispherically end-capped cylinder.

periodicity in the horizontal direction is produced by the periodicity of ribs on the structure surface.

Finally, we observe that the number of modes in Fig. 3 exceeds the number of modes in Fig. 4 because the contribution of the endcaps is included in the eigenfunction decomposition algorithm.

IV. SUMMARY

The similarities between the eigenfunction decomposition method and Fourier expansion on the cylindrical portion of the cylinder are observed for a framed cylindrical shell. The advantage of the eigenfunction expansion method is that the algorithm may be applied to a general structure with nonseparable geometry. We observe that it is useful in identifying the various waves present on the surface of the structure.

ACKNOWLEDGMENT

This work was supported by ONR.

[1] E. G. Williams, B. H. Houston, and J. A. Bucaro, ''Experimental investigation of the wave propagation on a point-driven, submerged capped cylinder using K-space analysis,'' J. Acoust. Soc. Am. **87**, 513–522 (1990).

[2] M. H. Marcus and B. H. Houston, ''The effect of internal point masses on the radiation of a ribbed cylindrical shell,'' J. Acoust. Soc. Am. **112**, 961–965 (2002).

[3] M. H. Marcus, B. H. Houston, and D. M. Photiadis, ''Wave localization on a submerged cylindrical shell with rib aperiodicity,'' J. Acoust. Soc. Am. **109**, 865–869 (2001).

[4] A. Sarkissian, ''Nearfield acoustic holography for an axisymmetric geometry: A new formulation,'' J. Acoust. Soc. Am. **88**, 961–966 (1990).

[5] A. Sarkissian, ''Nearfield acoustical holography,'' in *Acoustic Interactions with Submerged Elastic Structures*, edited by A. Guran, A. Bostrom, O. Leroy, and G. Maze (World Scientific, Singapore, 2002), pp. 239–276.

[6] H. Allik, R. Dees, S. Moore, and D. Pan, *SARA-2D User's Manual* (BBN Systems and Technologies, New London, CT, 1995).

Detection and localization of inclusions in plates using inversion of point actuated surface displacements

J. A. Bucaro,[a] A. J. Romano, and P. Abraham
The Naval Research Laboratory, Washington, DC 20375

S. Dey
SFA, Inc., Largo, Maryland 20774

(Received 19 June 2003; accepted for publication 10 October 2003)

A numerical simulation is carried out demonstrating the use of plate surface vibration measurements for detecting and locating inclusions within the structure. A finite element code is used to calculate normal surface displacement for both steel and mortar plates subjected to a monochromatic point force. The data is generated for the homogeneous plate and the identical plate within which exists a small rectangular inclusion. It is observed that when the elastic modulus of the inclusion is orders of magnitude lower than the base material, resonances of the inclusion produce large local displacements that are readily observed in the raw displacement data. For more modest moduli differences, there are no such directly observable effects. In this case, three inverse algorithms are used to process the displacement data. The first two are local inversion techniques that each yield a spatial map of the elastic modulus normalized by density. These algorithms successfully detect and localize the inclusion based on its modulus difference from that of the base plate. The third technique uses a form of the inhomogeneous equation of motion to obtain the induced force distribution connected with the inclusion. The spatial mapping of this force also successfully detects and localizes the inclusion. [DOI: 10.1121/1.1631936]

PACS numbers: 43.40.Le, 43.40.Dx, 43.60.Pt. [RLW] Pages: 201–206

I. INTRODUCTION

Interest in infrastructure health monitoring has continued to grow over the past decade. Monitoring techniques using the dynamic mechanical, acoustic, or ultrasonic response represent one of several major research focuses aimed at providing increased capabilities for monitoring fault development in structures. These mechanical approaches include modal analysis techniques (for example, see Wang *et al.*[1] and Kim and Stubbs[2]) typically based on resonance frequencies and/or associated mode shapes. One drawback of these modal approaches results from the fact that local changes in a structure often produce only very small changes in the global modal properties whereas unavoidable environmental changes can have a large impact. In addition, localization of the flaw is difficult using global mode responses.

In this paper we examine several new methods that also use the mechanical dynamic response but that, however, are able to detect *local* changes in the structural dynamics caused by the presence of a fault. The new methods, like modally based techniques, use measurements of surface displacement. In our case they are excited on a plate-like structure in response to a local, dynamic force. We consider three techniques. The first is direct observation of the spatial distribution of the surface displacement while the other two are considered inverse techniques. The first of the latter involves direct inversion of the equations of motion to obtain elastic moduli maps similar to that used by Romano *et al.*[3,4] to invert shear wave displacement in soft, tissue-like materials and the second a generalized force mapping using a form of the inhomogeneous equations of motion.

In Sec. II we present the mathematical foundation for the inverse techniques, and in Sec. III we apply the techniques to plate vibration data generated using finite element modeling.

II. MATHEMATICAL DESCRIPTION OF INVERSE TECHNIQUES

In the following section, we outline the mathematical basis for two inversion techniques that can be applied to the two-dimensional normal displacement data bases for the purposes of detecting and identifying subsurface flaws in the plate-like structures. We discuss these approaches in the context of Fig. 1. Here a solid plate structure lying in the x, y plane is forced dynamically at frequency ω at a point. By an appropriate method, the dynamic normal displacement $u_z(x,y,\omega)$ is measured over its surface with spatial resolution sufficient to resolve the smallest flexural wavelength. The flaw would then be detected and localized by applying the inversion algorithms discussed below to the measured displacement data $u_z(x,y,\omega)$.

A. Inversion for material elastic parameters

Romano *et al.*[3] have developed an inverse technique based on a variational formulation that is able to operate locally on interior displacement fields generated by a shear transducer in order to obtain three-dimensional maps of the elastic moduli/density ratios throughout the body. They subsequently demonstrated this inverse algorithm using *interior* displacements measured by magnetic resonance elastography

[a]Corresponding author. Electronic mail: joseph.bucaro@nrl.navy.mil

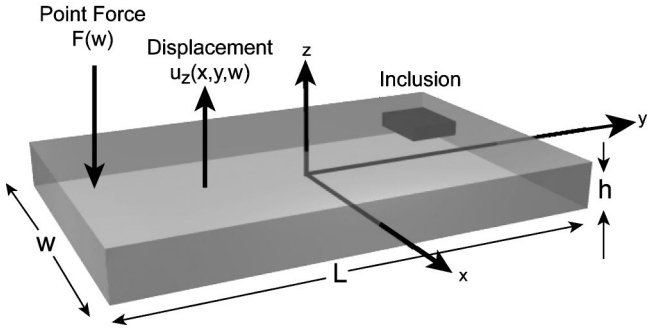


FIG. 1. Coordinate system and plate geometry used in numerical calculations for the normal displacement u_z at the plate surface in response to a normal point force F at frequency ω.

on human tissue phantoms.[4] Here, we attempt to apply a modified version of this approach to the *surface* displacement data $u_z(x,y,\omega)$ measured on the plates in response to a local dynamic point force.

As discussed by Romano *et al.*,[3,4] the equations of equilibrium[5] for general elastic media,

$$\sigma_{ij,j}[\vec{u}]+\rho\omega^2 u_i=0, \tag{1}$$

can be rewritten in the variational form

$$\int_{\Omega_m}(\sigma_{ij}[\vec{u}]e_{ij}[\vec{\nu}]-\rho\omega^2 u_i\nu_i)d\Omega=\int_{\Gamma_m}\nu_i t_i\,d\Gamma, \tag{2}$$

where σ_{ij} is the stress, $\sigma_{ij,j}$ denotes the partial derivative with respect to x_j, ε_{ij} the strain, i, j correspond to the x, y, z directions, ρ the density, ω the radial frequency, $\vec{u}$ the displacement, $\vec{\nu}$ smoothly varying virtual functions (described below), Ω_m the volume of the sampled displacement, Γ_m the corresponding circumscribing surface area, and t_i the components of surface traction. If we choose virtual functions such that they as well as their normal derivatives vanish on the surface Γ_m and expand the surface tractions in terms of the displacements and their derivatives, Eq. (2) reduces to

$$\int_{\Omega_m}u_i\sigma_{ij,j}[\vec{\nu}]d\Omega=-\int\rho\omega^2 u_i\nu_i\,d\Omega. \tag{3}$$

Upon dividing both sides of Eq. (3) by the density, this system may be used to solve for the material elastic parameters. Using the relation between σ_{ij} and ε_{ij} appropriate to the case for shear polarized in the z direction,[5] Romano *et al.*[3,4] reduce Eq. (3) to

$$\left(\frac{\mu}{\rho}\right)_m=\frac{\int\omega^2 u_z\nu_z\,d\Omega_m}{\int u_z\left[\dfrac{\partial^2\nu_z}{\partial x^2}+\dfrac{\partial^2\nu_z}{\partial y^2}\right]d\Omega_m}, \tag{4}$$

where μ is the shear modulus. Equation (4) is thus the inversion algorithm that operates on the measured displacement data $u_z(x,y,z)$ to obtain $\mu(x,y,z)/\rho(x,y,z)$.

We would like to obtain the equivalent expression for the flexural vibration case of interest here. Consider the equation of normal motion for a thin isotropic plate[6] of thickness h and density ρ, i.e.,

$$D\left(\frac{\partial^4 u_z}{\partial x^4}+\frac{\partial^4 u_z}{\partial x^2\partial y^2}+\frac{\partial^4 u_z}{\partial y^4}\right)-\rho\omega^2 h u_z=f(x,y,\omega), \tag{5}$$

where D is given in terms of the Young's modulus E and the Poisson ratio υ as $D=Eh^3/12(1-\nu^2)$ and f is a force applied normal to the surface. As a first attempt at inversion, consider the homogeneous form of Eq. (5) that would apply at points away from the imposed force. With $f=0$, Eq. (5) can be rewritten as

$$\frac{D}{\rho h}=\frac{\omega^2 u_z}{\nabla^4 u_z}. \tag{6}$$

However, the application of the fourth-order derivative to measured data (or to certain computed, descretized data such as by finite elements) is problematic since u_z and its derivatives as measured or computed are usually not continuous functions. As in Eq. (2), to avoid this issue Eq. (5) can be multiplied by the virtual function ν_z (again chosen such that the aforementioned quantities vanish on the surface), integrated over the volume Ω, and then used to solve for the modulus parameter D, with the result that

$$\frac{D}{\rho}=\frac{h\omega^2\int u_z\nu_z\,d\Omega}{\int\nu_z\nabla^4 u_z\,d\Omega}. \tag{7}$$

Making the crude but nonetheless applicable assumption that the surface displacements approximate those through the plate thickness, we obtain

$$\frac{Eh^2}{12\rho(1-\nu^2)}=\omega^2\frac{\int u_z\nu_z\,dS}{\int\nu_z\nabla^4 u_z\,dS}, \tag{8}$$

where the integrals are over the plate surface and where we now interpret the modulus on the left-hand side as the effective local bending modulus of the plate structure.

At this point, a lower-order version of Eq. (8) can be readily obtained by performing a fourfold integration by parts on the denominator and using the fact that the virtual functions and their first derivatives vanish on the surface. This results in an expression for the denominator on the right-hand side for which there exists no derivative of $u_z(x,y)$ higher than the first. This is important when one considers the impact of the higher-order derivatives acting on noisy data.

Carrying out the integration by parts of the denominator in Eq. (8) four times yields the following expression for flexure analogous to Eq. (4) for shear:

$$\frac{Eh^2}{12(1-\nu^2)\rho}=\omega^2\frac{\int u_z\upsilon_z\,dx\,dy}{I}. \tag{9}$$

Here the denominator is now given by

$$I=\int_y\left[\left.\frac{\partial^2\nu_z}{\partial x^2}\frac{\partial u_z}{\partial x}\right|_x-\left.\frac{\partial^3\nu_z}{\partial x^3}u_z\right|_x+\int_x u_z\frac{\partial^4\nu_z}{\partial x^4}dx\right]dy$$
$$+2\int_x\int_y u_z\frac{\partial^4\nu_z}{\partial x^2\partial y^2}dx\,dy+\int_x\left[\left.\frac{\partial^2\nu_z}{\partial y^2}\frac{\partial u_z}{\partial y}\right|_y-\left.\frac{\partial^3\nu_z}{\partial y^3}u_z\right|_y\right.$$
$$\left.+\int_y u_z\frac{\partial^4\nu_z}{\partial y^4}dy\right]dx. \tag{10}$$

Expression (9) together with (10) termed the "weak" flexural formulation has the advantage that no higher than first-order spatial derivatives operate on the measured data

$u_z(x,y)$. As can be seen, we are now able to obtain the ratio of the appropriate elastic parameter $D/(\rho h)$ by a fairly well-behaved operator acting on the measured displacement field u_z. Accordingly, if the flaw has an appreciable impact on the local bending stiffness, processing the measured surface displacements according to Eqs. (9) and (10) can serve to detect and localize the fault.

B. Generalized force mapping

The concept we call generalized force mapping results from a straightforward application of the inhomogeneous equation of motion in Eq. (5) that we rewrite here:

$$D\nabla^4 u_z - \rho\omega^2 h u_z = f_z. \tag{11}$$

The generalized force mapping technique we suggest here essentially uses the measured displacement field, $u_z(x,y,\omega)$, and the known parameters of the homogeneous plate, i.e., $D/(\rho h)$, and computes the sum of the two terms on the left-hand side of Eq. (11). For a homogeneous plate, away from the applied force, these terms add identically to zero. For the plate with the inclusion, we interpret a nonzero result for the left-hand side (away from the known position of the applied force) as a generalized force, $f_z(x,y,\omega)$, which exists in the affected region as a consequence of the presence of the flaw. The mapping of these body forces that become active in the vicinity of the inclusion when the plate is excited then yields an induced force distribution that serves to detect and locate the flaw. Again, we can capitalize on the development leading up to Eq. (9) and the expression for I in Eq. (10) to reduce the order of the derivative acting on u_z. Equation (11) then becomes

$$ID/(\rho h) - \omega^2 \int u_z v_z \, dx\, dy = G, \tag{12}$$

where now $G = \int f_z \nu_z \, dx\, dy$ is the force parameter that serves to detect and localize the fault.

III. NUMERICAL RESULTS

We focus our study and demonstration of the techniques on a solid homogeneous plate of dimensions length $=60$ cm, width$=30$ cm, and thickness$=2.54$ cm with and without an internal inclusion. We consider two different sets of material parameters for the homogeneous plate, including the density, ρ, the Young's modulus, E, and the Poisson ratio, υ. The two sets correspond to mortar and to steel. The flaw is in the form of a rectangular inclusion (see Fig. 1) of length $=2$ cm, width$=1.5$ cm, and thickness$=0.5$ cm (centered at $x=-8.7$ cm, $y=19$ cm, $z=1.27$ cm) and with the following material parameters: density$=\rho$, Young's modulus$=\beta_f E$, and Poisson ratio$=\upsilon$.

A. Finite element database

We used a finite element structural dynamics code to compute the surface displacements of these structures resulting from a normal point force applied to the plates near their lower left corner (see Fig. 1). The responses were computed for several frequencies covering a band that we believe would be easy to implement in practice, both from a force application point of view and from a scanned surface displacement measurement point of view. The frequencies chosen are 10, 20, 25, and 50 kHz. For each case we compute the normal surface displacement $u_z(x,y,\omega)$ on a rectangular grid with a spacing of 0.25 cm.

B. Direct observation of surface displacement

In this section, we discuss the surface displacement databases, first to describe the nature of the data that will be operated upon by the inverse algorithms, and also to show that there are circumstances in which a direct observation of this data can be used to detect the flaw.

As a representative example, we show in Fig. 2 the displacement data computed for the mortar plate. The upper and lower rows are for the 10 and 25 kHz cases, respectively. One can see the expected development of a modal response of the plate structure. The first column corresponds to the plate with no inclusion, while the second, third, and fourth columns correspond to the plate with the flaw for the case $\beta_f=0.05$, 0.01, and 0.001, respectively, i.e., the cases where the flaw has a Young's modulus 5%, 1%, and 0.1% of the modulus of the mortar itself. For the $\beta_f=0.05$ case, the presence of the inclusion is not indicated in these direct observations. Accordingly, data of this type will provide a good test of the ability of the inversion algorithms to detect and locate the flaws, and this will be presented in the next sections for the steel plate case.

Consider the case shown in the right column of data. This is for the extreme case in which $\beta_f=10^{-3}$. For the mortar plate, this condition might be representative of the formation of an air layer, or delamination/detachment within the structure. For this extreme case (and to a lesser degree for 0.01 as well), one can see that direct observation of the displacement data clearly indicates the presence and location of the internal flaw. This is not surprising and is consistent with the large displacements that would be generated locally at the various resonance frequencies of this very soft cavity. In fact, one expects a number of resonances in our frequency range, including the first three lengthwise, the first two widthwise, and the first thicknesswise resonances. We point out that we have recently successfully used this direct observation technique and delamination resonance to identify mortar fault layers in mural pilasters in the United States Capitol Building[7] using surface displacement data we obtained with a scanning laser Doppler vibrometer.

C. Inversion for elastic parameters and generalized force mapping

We now consider the application of the inverse algorithms in order to evaluate their efficacy in detecting and locating the internal inclusion for the more typical cases in which the displacement data does not directly indicate the presence of the flaw. For such a case, we show the results of applying both algorithms to the steel plate displacement ''data'' obtained at 20 kHz in Fig. 3. The inclusion has a Young's modulus parameter β_f of 0.05.

In the leftmost column of Fig. 3 is shown the displacement ''data'' for both the homogeneous plate and the plate

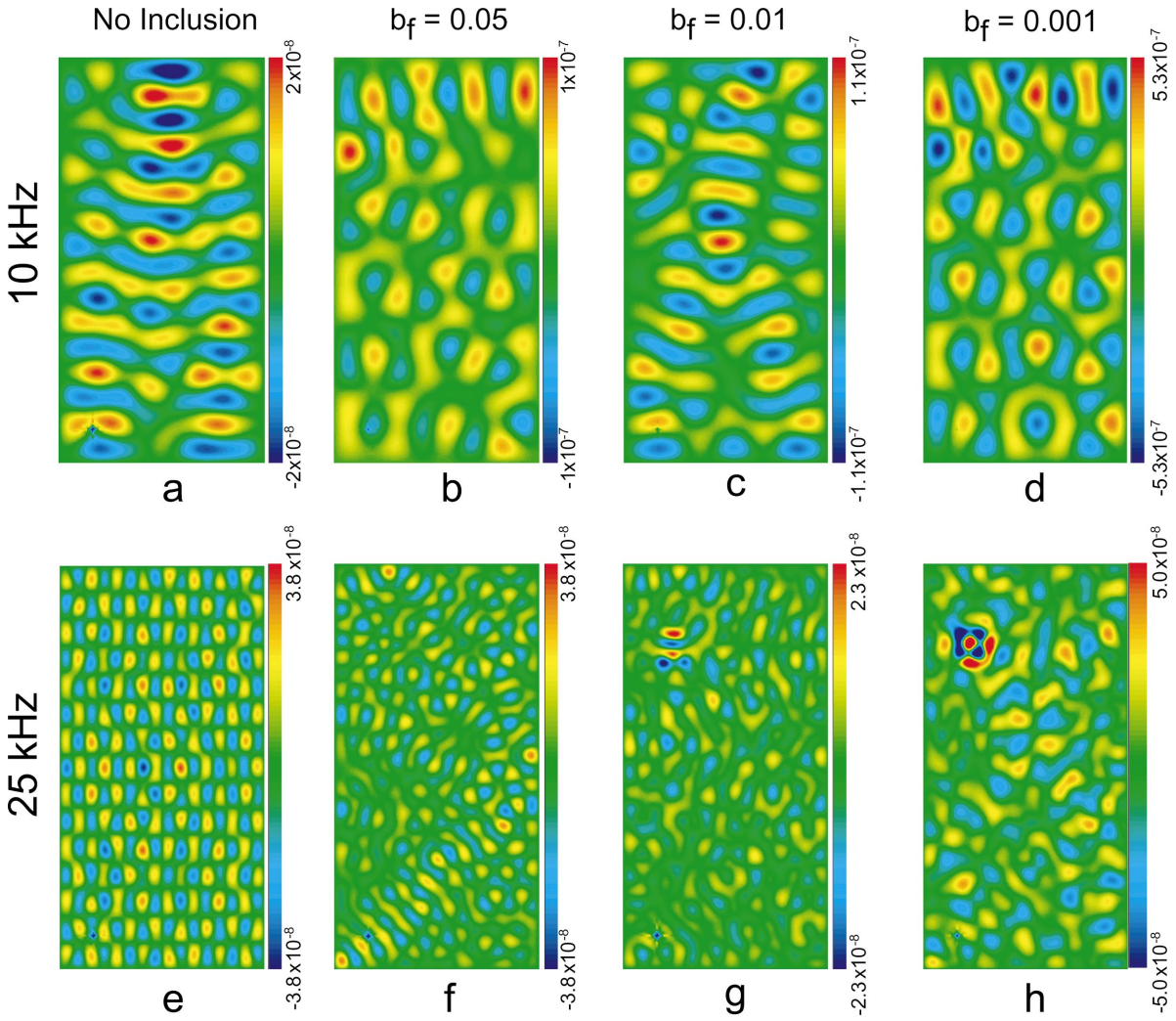


FIG. 2. Surface displacements calculated for the mortar plate at 10 and 25 kHz for no inclusion (a), (e), an inclusion with Young's modulus of 0.05 E (b), (f), 0.01 E (c), (g), and 0.001 E (d), (h). The color scale gives displacement amplitudes in arbitrary units.

with the inclusion. Clearly there does not appear to be any definitive indication of the location of the flaw. In the next column is shown the result of applying the inverse algorithm specified by Eqs. (9) and (10). In both plates, one can see a somewhat regular pattern throughout the elastic modulus map. This pattern, which more or less corresponds to the modal response of the plate at this frequency, is an artifact of the inversion algorithm resulting from the zero displacement levels (i.e., $u_z=0$) which exist at the nodal lines of the modes. The inversion algorithm is not conditioned to properly handle these zero levels. This anamolous structure in the elastic moduli maps can be eliminated by using broadband data so that the nodal lines for any particular mode do not lead to zero displacement levels anywhere across the plate. This must be done under the proviso that the chosen band not be so large that dispersion becomes important.

Ignoring these nodal line artifacts, one can clearly see the rectangular flaw in the upper left-hand corner of the lower modulus map. Further, the color map gives a value for the ratio of the modulus $D/(h\rho)$ for the homogeneous region versus the flawed region of about 3.8. For a 2.54 cm plate of material of modulus E with a 0.5 cm layer at its midplane of modulus $0.05E$, we estimate that the effective modulus would be $0.32D/(h\rho)$. Thus, we expect from the inversion map that the ratio of moduli away from the flaw compared to that over the flaw would be about 3, which is close to that actually found.

To demonstrate the importance of developing the ''weak'' formulation of the inversion algorithm, i.e., Eqs. (9) and (10) versus Eq. (6), we apply the latter to the displacement data and show the results in the third column of Fig. 3. As can be seen, in addition to the pattern associated with the nodal lines, there exists a finer scaled pattern as well. This is an artifact caused by an application of the fourth-order derivative on the descretized (and therefore ''noisy'') displacement data generated by the finite element calculations. As discussed at the end of Sec. II A, this effect is eliminated in Eqs. (9) and (10), where the higher-order derivatives operate on the smooth virtual functions rather than the ''measured'' data. In this particular case, notwithstanding these dominant spurious features, the inclusion is still readily observable.

Next, we consider the application of the generalized force mapping algorithm. Shown in the right column of Fig. 3 are the results for this algorithm [Eq. (12)] for the homogeneous plate (upper figure) and for the plate with the inclusion (lower figure). As can be seen, the generalized force

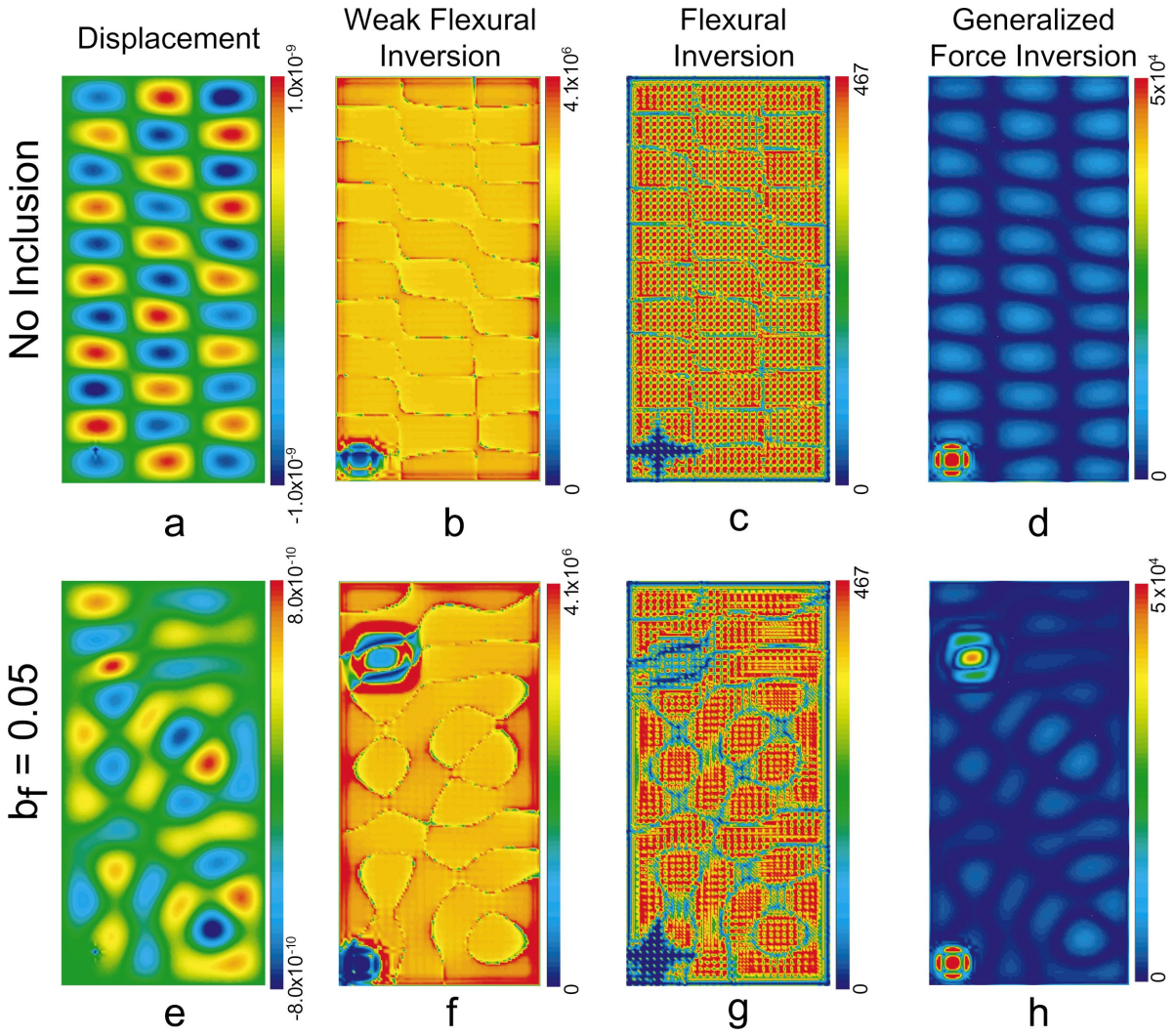


FIG. 3. Results for the steel plate excited by a point force at 20 kHz: Displacement numerical data for no inclusion (a) and for inclusion with Young's modulus 0.05 E (e); the application of weak flexural inversion algorithm [Eqs. (9), (10)] for no inclusion (b) and for inclusion (f); the application of the flexural inversion algorithm [Eq. (6)] for no inclusion (c) and for inclusion (g); the application of a generalized force inversion algorithm [Eq. (12)] for no inclusion (d) and for inclusion (h).

map does detect and locate the inclusion (and the applied force as well). Compared to the elastic moduli inversion algorithm, this algorithm is not as affected by the presence of nodal lines.

IV. CONCLUDING REMARKS

The results shown above are intended to illustrate and demonstrate under very simplifying conditions what we consider to be an interesting and promising methodology for internal flaw detection and identification in plate-like structures. The approach utilizes dynamic surface displacement data generated in response to a point actuator that, in general, can be obtained in a number of ways. One of the more obvious methods, and one we have used often, is that involving a scanning laser Doppler vibrometer. What we have not considered in this paper are complicating issues relevant to specific applications such as measurement noise, spatial sampling requirements, material damping, etc. We are, in fact, carrying out further studies to quantify the impact of these effects, and when completed those results will be presented in a future article. However, these complications do not, in general, obviate the merit to applying these approaches to real structures. In fact, we have already successfully employed versions of these techniques on plate structures in the laboratory and on pilaster walls in the U.S. Capitol Building as part of an art restoration project.[7]

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research.

[1] M. L. Wang, F. L. Xu, and G. M. Lloyd, "A systematic numerical analysis of the damage index method used for bridge diagnostics," *Proceedings of the SPIE*, Smart Structures and Materials, 2000, Vol. 3988, pp. 154–164.
[2] J. T. Kim and N. Stubbs, "Model-uncertainty impact and damage-detection accuracy in plate girder," J. Struct. Eng. **121**, 1409–1417 (1995).
[3] A. J. Romano, J. J. Shirron, and J. A. Bucaro, "On the non-invasive determination of material parameters from a knowledge of elastic displacements: theory and numerical simulation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 751–759 (1998).

[4] A. J. Romano, J. A. Bucaro, R. L. Ehman, and J. J. Shirron, ''Evaluation of a material parameter extraction algorithm using MRI-based displacement measurements,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1575–1581 (2000).
[5] B. A. Auld, Acoustic Fields and Waves in Solids (Wiley, New York, 1973), Vol. 1, p. 68.
[6] M. C. Junger and D. Feit, ''Sound, structures, and their interactions,'' Acoust. Soc. Am. 210 (1993).
[7] J. A. Bucaro, J. Vignola, B. H. Houston, and A. J. Romano, ''Preliminary observations regarding LDV scans of panels excited by broadband actuators at the US Capitol,'' NRL Memorandum Report #8685, 18 January 2003.

Hybrid near-field acoustic holography

Sean F. Wu[a)]
Department of Mechanical Engineering, Wayne State University, Detroit, Michigan 48202

(Received 21 February 2003; accepted for publication 13 October 2003)

Hybrid near-field acoustical holography (NAH) is developed for reconstructing acoustic radiation from an arbitrary object in a cost-effective manner. This hybrid NAH is derived from a modified Helmholtz equation least squares (HELS) formula that expands the acoustic pressure in terms of outgoing and incoming waves. The expansion coefficients are determined by solving an overdetermined linear system of equations obtained by matching the assumed-form solution to measured acoustic pressures through the least squares. Measurements are taken over a conformal surface around a source at close range so that the evanescent waves can be captured. Next, the modified HELS is utilized to regenerate as much acoustic pressures on the conformal surface as necessary and take them as input to the Helmholtz integral formulation implemented numerically by boundary element method (BEM). The acoustic pressures and normal velocities on the source surface are reconstructed by using a modified Tikhnov regularization (TR) with its regularization parameter determined by generalized cross validation (GCV) method. Results demonstrate that this hybrid NAH combines the advantages of HELS and inverse BEM. This is because a majority of the input data are regenerated but not measured, thus the efficiency of reconstruction is greatly enhanced. Meanwhile, the accuracy of reconstruction is ensured by the Helmholtz integral theory and modified TR together with GCV method, provided that HELS converges fast enough on the measurement surface. Numerical examples of reconstructing acoustic quantities on the surface of a simplified engine block are demonstrated. [Work supported by NSF.] [DOI: 10.1121/1.1631415]

PACS numbers: 43.40.Sk, 43.20.Ye, 43.60.Pt [EGW] Pages: 207–217

I. INTRODUCTION

Since its first publication in 1980 near-field acoustical holography (NAH)[1–4] has shown great promise in reconstructing acoustic radiation from vibrating structures based on the acoustic pressures measured on a hologram plane. Traditional NAH is carried out via Fourier acoustics,[5] which transforms the acoustic pressure and normal velocity from the time domain to frequency domain, followed by another transform to the wavenumber domain. Once the acoustic quantities are specified in the wavenumber domain, they are transformed back to the frequency and time domains. This Fourier acoustics is suitable for separable geometry such as an infinite plane, an infinite cylinder, and a sphere in a free field. Generalized NAH can be derived using the Green's function that satisfies the homogeneous Dirichlet or Neumann condition on a hologram plane.[6–8] Once this Green's function is specified, the entire acoustic field can be reconstructed. However, the desired Green's function for an arbitrary surface is extremely difficult to derive. Thus other methodologies are sought. One such method is the Helmholtz integral formulation implemented numerically through boundary element method (BEM), which is used by Gardner and Bernhard[9] to characterize acoustic sources in a region enclosed by an arbitrarily shaped surface. Veronesi and Maynard[10] present an inverse BEM (IBEM) for reconstructing acoustic radiation from an arbitrary structure. Since then many papers on IBEM have emerged.[11–15]

The main advantages of IBEM are its applicability to an arbitrary surface and flexibility in taking acoustic pressure measurements. However, the surface acoustic quantities are described by spatial discretization, which requires a minimum number of nodes per wavelength in a surface mesh to avoid distortions. Accordingly, one must take a comparable though not necessarily equal number of measurements to solve the surface acoustic quantities. For a complex structure such as a vehicle vibrating at low-to-mid frequencies, the number of nodes required in the mesh can be huge. As a result, the reconstruction process can be very complex and time consuming.

Recently, another technique was developed that expresses acoustic pressure as an expansion of the particular solutions to the Helmholtz equation. The coefficients of expansion are obtained by matching the assumed-form solution to the measured acoustic pressures through least squares, which is known as Helmholtz equation least squares (HELS) method.[16,17] Since HELS solves the Helmholtz equation directly, it is immune to the nonuniqueness difficulty[18,19] inherent in IBEM.

The main advantages of HELS are its simplicity in mathematical formulation, efficiency in numerical computations, and flexibility in applications. HELS has been successfully used to reconstruct acoustic radiation from arbitrary structures in exterior[20,21] and interior[22,23] regions. Moreover, test results demonstrate that HELS enables one to reconstruct the acoustic fields with relatively few measurements and allows piece-wise reconstruction, just like patch NAH.[24]

However, test results also indicate that HELS is not ideal for a highly irregular structure due to a slow convergence of

a)Electronic mail: swu@me1.eng.wayne.edu

the expansion solution. To circumvent this difficulty, a combined HELS known as CHELS method[25] is developed, which enables one to tackle arbitrary geometry with relatively few measurements but at the expense of reconstruction accuracy.

All these NAH techniques are limited to objects vibrating in an unbounded fluid medium. In practice, most vibrating structures are mounted on solid foundations. A typical example is the analysis of noise radiation from a vehicle stationed on the chassis dynamometers inside a semianechoic chamber. For safety and durability concerns, the surfaces of the test chamber cannot be made as acoustically absorptive as they should be. As a result, the field acoustic pressures consist of both direct and reflected waves, making NAH application and analysis very difficult. To date, vehicle noise is still analyzed by measuring transfer functions between a source and receiver, or by sweeping an intensity probe over the vehicle surface at very close range. The information thus obtained is limited to measurement locations and the process is very time consuming.

The objective of this paper is to develop hybrid NAH to reconstruct acoustic radiation from an arbitrary object in confined or free space in a cost-effective manner. This hybrid NAH is derived from a modified HELS, which expands the acoustic pressure in terms of outgoing and incoming waves, and combines it with IBEM plus regularization. Since this hybrid NAH allows regeneration of the acoustic pressure on a measurement surface, both the accuracy and efficiency of reconstruction are enhanced when HELS converges fast on the measurement surface.

II. BASIC THEORY

Consider acoustic radiation from an arbitrary object vibrating at a constant frequency ω in an unbounded fluid medium. In the original HELS the acoustic pressure is expressed as

$$p(\mathbf{x};\omega)=\sum_{j=1}^{J}\Psi_j^{(1)}(\mathbf{x};\omega)C_j(\omega), \tag{1}$$

where $\Psi_j^{(1)}$ is the jth particular solution to the Helmholtz equation in any coordinate system. For example, using the spherical coordinates we can write $\Psi_j^{(1)}$ as

$$\Psi_j^{(1)}\equiv\Psi_{nl}^{(1)}(r,\theta,\phi;\omega)=h_n^{(1)}(kr)Y_n^l(\theta,\phi), \tag{2}$$

where $h_n^{(1)}(kr)$ imply the spherical Hankel functions of order n of the first kind, k is the acoustic wavenumber, $Y_n^l(\theta,\phi)$ are the spherical harmonics, and the indices j, n, and l in (2) are related via $j=n^2+n+l+1$ with n starting from 0 to N and l varying from $-n$ to n. Thus, for each n and l we have $j=1$ to J, where $J=(N+1)^2$ indicates the maximum number of expansion functions. The expansion coefficients C_j are determined by matching the assumed-form solution (1) to the acoustic pressure $\tilde{p}(\mathbf{x}_m^\Gamma;\omega)$ measured on a conformal surface Γ around the source,

$$p_J(\mathbf{x}_m^\Gamma;\omega)=\tilde{p}(\mathbf{x}_m^\Gamma;\omega),\quad \mathbf{x}_m^\Gamma\in\Gamma,\quad m=1\ \text{ to }\ \tilde{M}\quad(\tilde{M}>J). \tag{3a}$$

The overdetermined system of equations (3a) is equivalent to a least-squares problem,

$$\min_{C_1,C_2,\cdots,C_J}\|p_J(\mathbf{x}_m^\Gamma;\omega)-\tilde{p}(\mathbf{x}_m^\Gamma;\omega)\|_2^2. \tag{3b}$$

Note that here we utilize a portion of the measured data to determine the expansion coefficients, $\mathbf{x}_m^\Gamma\in\Gamma$, $m=1$ to $M_0\,(J<M_0<\tilde{M})$. This is because Eq. (3) can only provide a best fit of the reconstructed acoustic pressures on Γ, but not necessarily elsewhere, especially when projecting the acoustic pressure back to a source surface. In fact, without regularization the reconstruction accuracy on a source surface increases with the expansion functions at first and then deteriorates quickly for a given set of measurements. This behavior is caused by the fact that the evanescent waves that dominate the acoustic near field are lost in the measured data. Consequently, when the high-order terms are used in reconstruction, what we get is nothing but an amplified noisy signal.

An optimal number of the expansion functions J_{op} with respect to the measured data on Γ can be defined as the smallest number of terms that yields the minimum errors in Eq. (3) with respect to all measured data for a given tolerance.[17] This is equivalent to finding a low-pass filter to eliminate the small singular values that drop below the ambient noise level. Methods that have been tested to yield J_{op} include an iteration scheme[17] and a constrained minimization.[26] The former captures a J_{op} that corresponds to the minimum mean-squared errors for any given set of input data, while the latter tends to yield a slightly lower value J_{op}, known as oversmoothing. These methods all attempt to expand acoustic fields in terms of the dominant acoustic modes, just like truncated singular value decomposition (SVD):

$$\min_J\|p_J(\mathbf{x}_m^\Gamma;\omega)-\tilde{p}(\mathbf{x}_m^\Gamma;\omega)\|_2^2,\quad \mathbf{x}_m^\Gamma\in\Gamma,\quad i=1\ \text{ to }\ \tilde{M}. \tag{4}$$

Once J_{op} is specified, the acoustic pressure anywhere can be reconstructed using Eq. (1).

It is emphasized that Eq. (1) is valid in the exterior region including an arbitrarily shaped source surface S. The completeness of an expansion in terms of the particular solutions to the Helmholtz equation was first demonstrated by Vekua over half a century ago.[27] Unfortunately, his paper did not receive much attention until Millar[28,29] brought out Vekua's proof in his studies of completeness of certain families of functions 20 years later. Isakov and Wu[30] demonstrate a rigorous mathematical justification of using Eq. (1) together with the least squares and quasisolution methods for reconstructing acoustic radiation from an arbitrary object. In particular, they prove that a radiating solution to the Helmholtz equation outside a sufficiently smooth, convex, and bounded domain can be approximated by a family of special solutions.[30] Using Eq. (1) and the conditional stability estimates in the Cauchy problem for an elliptic equation, Isakov and Wu prove that these special solutions are bounded on and outside a source surface and converge to the exact solution, provided that they converge to the exact solution on a

measurement surface. The same conclusions also hold for reconstructing the acoustic field in an interior region.[30]

III. HYBRID NAH FORMULATIONS

A. Modified HELS

In order to enhance the efficiency of reconstruction of acoustic radiation from arbitrarily shaped source geometry, we propose to rewrite the near-field acoustic pressure as

$$p(\mathbf{x};\omega)=\sum_{j=1}^{J}\Psi_j^{(1)}(\mathbf{x};\omega)C_j(\omega)+\sum_{j=1}^{J}\Psi_j^{(2)}(\mathbf{x};\omega)D_j(\omega), \tag{5}$$

where $\Psi_j^{(1)}$ is given by Eq. (2) and $\Psi_j^{(2)}\equiv\Psi_{nl}^{(2)}(r,\theta,\phi)=h_n^{(2)}(kr)Y_n^l(\theta,\phi)$, if spherical coordinates are used, with $h_n^{(2)}(kr)$ being the spherical Hankel functions of order n of the second kind.

Physically, the two terms on the right side of Eq. (5) represent the outgoing and incoming spherical waves, respectively. To facilitate derivations of hybrid NAH formulations, we rewrite these expansion functions as a matrix and determine the expansion coefficients by solving an overdetermined linear system of equations obtained by matching the assumed-form solution (5) to the acoustic pressures measured on Γ through least squares.

$$\tilde{\mathbf{p}}(\mathbf{x}_m^{\Gamma};\omega)=\tilde{\mathbf{A}}(\mathbf{x}_m^{\Gamma};\omega)\hat{\mathbf{C}}(\omega),\quad \mathbf{x}_m^{\Gamma}\in\Gamma,\quad m=1\ \text{ to }\ \tilde{M}, \tag{6}$$

where $\tilde{\mathbf{p}}(\mathbf{x}_m^{\Gamma};\omega)$ implies a column vector of acoustic pressures on Γ, and $\tilde{\mathbf{A}}$ and $\hat{\mathbf{C}}$ are given by

$$\tilde{\mathbf{A}}=\begin{bmatrix}\Psi_{11}^{(1)} & \Psi_{11}^{(2)} & \Psi_{12}^{(1)} & \Psi_{12}^{(2)} & \cdots & \Psi_{1J}^{(1)} & \Psi_{1J}^{(2)}\\ \Psi_{21}^{(1)} & \Psi_{21}^{(2)} & \Psi_{22}^{(1)} & \Psi_{22}^{(2)} & \cdots & \Psi_{2J}^{(1)} & \Psi_{2J}^{(2)}\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ \Psi_{J1}^{(1)} & \Psi_{J1}^{(2)} & \Psi_{J2}^{(1)} & \Psi_{J2}^{(2)} & \cdots & \Psi_{JJ}^{(1)} & \Psi_{JJ}^{(2)}\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ \Psi_{\tilde{M}1}^{(1)} & \Psi_{\tilde{M}1}^{(2)} & \Psi_{\tilde{M}2}^{(1)} & \Psi_{\tilde{M}2}^{(2)} & \cdots & \Psi_{\tilde{M}J}^{(1)} & \Psi_{\tilde{M}J}^{(2)}\end{bmatrix}_{\tilde{M}\times 2J} \tag{7}$$

$$\hat{\mathbf{C}}=\begin{Bmatrix}C_1\\ D_1\\ C_2\\ D_2\\ \vdots\\ C_J\\ D_J\end{Bmatrix}_{2J\times 1},\quad \tilde{M}>2J.$$

From linear algebra we learn that Eq. (6) has a unique solution if and only if the matrix $\tilde{\mathbf{A}}$ has a full rank. In reconstruction, the problem is ill posed and the matrix is ill conditioned. In addition, the system may be underdetermined, namely, having less number of rows than that of columns in Eq. (6). Physically, this implies that there are fewer measurements than the unknown variables. Mathematically, we can still obtain a reasonably accurate solution using regularization, just as we do to an overdetermined system, provided that the right regularization parameter is selected. However, this is case dependent. In other words, it depends on how ill posed a problem is. The more ill posed a problem or more ill conditioned a matrix is, the sharper the drop from the largest singular value to the smallest one in the SVD spectrum is, and the more likely it is to find the right regularization parameter to eliminate the small singular values and to obtain a satisfactory reconstruction out of an underdetermined system.

On the other hand, the less ill conditioned a matrix is, the smoother the SVD spectrum is, and the less likely it is to find a proper regularization parameter to obtain a good reconstruction. This is because for weakly ill-posed problems, an underdetermined system resulting from taking fewer measurements than unknown variables means losing vital information in the input. Under this condition, there is no way of getting an accurate reconstruction and the reconstructed images may be distorted.[31] Since we do not know how ill conditioned a matrix is before measurements are taken, and since lacking information is one major cause of ill posedness in NAH, we will be well advised to use common sense and take sufficient measurements in practice.

Equation (5) can be employed to reconstruct acoustic radiation from a source in confined space. This can be done by taking measurements over two conformal surfaces Γ_1 and Γ_2 around a source at close range, with Γ_2 being inside Γ_1 by a small distance Δ. For example, we can take M_1 measurements on Γ_1 to solve the expansion coefficients $\hat{C}_j$, $j=1$ to $2J$, and optimize the number of expansion functions J_{op} by minimizing the reconstruction errors with respect to some additional M_2 measurements taken on Γ_2:

$$\min_{\hat{C}_1,\hat{C}_2,\cdots,\hat{C}_J}\left\|p_J(\mathbf{x}_m^{\Gamma_1};\omega)-\tilde{p}(\mathbf{x}_m^{\Gamma_1};\omega)\right\|_2,$$

$$\mathbf{x}_m^{\Gamma_1}\in\Gamma_1,\quad m=1\ \text{ to }\ 2J, \tag{8a}$$

$$\min_J\left\|p_J(\mathbf{x}_m^{\Gamma_2};\omega)-\tilde{p}(\mathbf{x}_m^{\Gamma_2};\omega)\right\|_2,$$

$$\mathbf{x}_m^{\Gamma_2}\in\Gamma_2,\quad m=1\ \text{ to }\ M_2. \tag{8b}$$

The reason for taking measurements over two concentric conformal surfaces for a source in confined space is to ensure that we acquire not only acoustic pressures but also their gradients so as to discern the directions of wave propagations. This is especially important for a reactive acoustic field, namely, the confining surfaces are highly reflective. Examples of such are shown in visualizing acoustic fields bounded by hard surfaces, for which reconstruction cannot possibly be done correctly without the use of double layers of measurements.[32]

It is emphasized that the modified HELS formula (5) can also be utilized to reconstruct acoustic radiation from an arbitrarily shaped structure vibrating at a constant frequency in a free field. Under this condition, it is sufficient to take measurements over a single conformal surface at very close range to the source. Taking measurements on two conformal surfaces only prolongs the reconstruction process with no apparent benefit of improving the reconstruction accuracy.

To enhance the efficiency and accuracy of reconstruction, we propose to combine Eq. (5) and IBEM together with regularization.

B. The Helmholtz integral theory

It is well known that the acoustic pressure at any point $\mathbf{x}$ exterior to a source surface S is expressible as the Fredholm integral equation of the first kind,[33]

$$p(\mathbf{x};\omega)=\frac{1}{4\pi}\int_S\int\left[p(\mathbf{x}_S;\omega)\frac{\partial G(\mathbf{x}|\mathbf{x}_S)}{\partial n_S}-G(\mathbf{x}|\mathbf{x}_S)\frac{\partial p(\mathbf{x}_S;\omega)}{\partial n_S}\right]dS, \tag{9}$$

where $G(\mathbf{x}|\mathbf{x}_S)=e^{ikR}/R$ is the free-space Green's function, $R=|\mathbf{x}-\mathbf{x}_S|$ is the distance between a field point $\mathbf{x}$ and a surface point $\mathbf{x}_S$ on S, and $\partial G/\partial n_S$ represents the normal derivative of G with respect to the surface unit normal at $\mathbf{x}_S$. The normal derivative of the surface acoustic pressure is related to the normal component of surface velocity v by the Euler's equation $\partial p/\partial n_S=i\omega\rho_0 v$. These surface acoustic quantities cannot be specified independently, since they are related via the Helmholtz integral equation that can be obtained by taking the limit as $\mathbf{x}\to\mathbf{x}_{S'}$. in Eq. (9),

$$C(\mathbf{x}_{S'})p(\mathbf{x}_{S'};\omega)=\frac{1}{2\pi}\int_S\int\left[p(\mathbf{x}_S;\omega)\frac{\partial G(\mathbf{x}_{S'}|\mathbf{x}_S)}{\partial n_S}-G(\mathbf{x}_{S'}|\mathbf{x}_S)\frac{\partial p(\mathbf{x}_S;\omega)}{\partial n_S}\right]dS, \tag{10}$$

where $C(\mathbf{x}_{S'})$ is constant and depends on local geometry of the source surface S that is given by

$$C(\mathbf{x}_{S'})=1+\frac{1}{4\pi}\int_S\int\partial/\partial n_S(1/R_S)dS, \tag{11}$$

where $R_S=|\mathbf{x}_{S'}-\mathbf{x}_S|$ is the distance between two surface points.

For an arbitrary surface there are no close-form solutions to Eqs. (9) and (10). Therefore they are solved numerically. Suppose that the surface is discretized into elements with N_0 nodes. The surface acoustic quantities are obtained using BEM by taking M measurements ($M\geqslant N_0$):

$$\begin{aligned}&\mathbf{T}_p(\mathbf{x}_m^\Gamma|\mathbf{x}_S;\omega)\mathbf{p}(\mathbf{x}_S;\omega)=\mathbf{p}(\mathbf{x}_m^\Gamma;\omega)\\ &\text{and}\\ &\mathbf{T}_v(\mathbf{x}_m^\Gamma|\mathbf{x}_S;\omega)\mathbf{v}(\mathbf{x}_S;\omega)=\mathbf{p}(\mathbf{x}_m^\Gamma;\omega),\end{aligned} \tag{12}$$

where $\mathbf{p}(\mathbf{x}_m^\Gamma;\omega)$ and $\mathbf{p}(\mathbf{x}_S;\omega)$ are column vectors of acoustic pressures on the measurement and source surfaces Γ and S, respectively, $\mathbf{T}_p$ and $\mathbf{T}_v$ represent the transfer matrices that correlate $\mathbf{p}(\mathbf{x}_m^\Gamma;\omega)$ to $\mathbf{p}(\mathbf{x}_S;\omega)$ and $\mathbf{v}(\mathbf{x}_S;\omega)$, respectively,

$$\mathbf{T}_p(\mathbf{x}_m^\Gamma|\mathbf{x}_S;\omega)=(4\pi)^{-1}[\mathbf{D}+\mathbf{M}\mathbf{M}_S^{-1}(2\pi\mathbf{I}-\mathbf{D}_S)], \tag{13a}$$

$$\mathbf{T}_v(\mathbf{x}_m^\Gamma|\mathbf{x}_S;\omega)=-(4\pi)^{-1}[\mathbf{D}(2\pi\mathbf{I}-\mathbf{D}_S)^{-1}\mathbf{M}_S+\mathbf{M}], \tag{13b}$$

where $\mathbf{M}$ and $\mathbf{D}$ imply the effects of monopoles and dipoles on a field point, respectively, and $\mathbf{M}_S$ and $\mathbf{D}_S$ represent the effects of monopoles and dipoles on a surface point, respectively.[33]

Since the surface acoustic quantities are described by spatial discretization, we must have a minimum number of nodes per wavelength to avoid distortions and take a comparable amount of measurements to reconstruct the acoustic quantities on these nodes. Such an approach may be unrealistic in practice because for a complex structure such as an automobile vibrating at low-to-mid frequencies, the discrete nodes and corresponding measurements may be excessive and the reconstruction process can be extremely complex and time consuming.

On the other hand, the modified HELS formulation (5) together with (8) can guarantee an accurate reconstruction of the acoustic pressures on the measurement surface Γ with relatively few measurements, but cannot guarantee simultaneously the accuracy of reconstructing acoustic pressures and normal velocities on an arbitrary source surface S, unless additional regularization is carried out and the proper regularization parameter is obtained.

In order to enhance the reconstruction accuracy and efficiency for the acoustic quantities on an arbitrary source surface, we combine Eqs. (5) and (10). Suppose that the source surface is discretized into elements with N_0 nodes, which satisfies a certain minimum number of nodes per wavelength requirement. Assume that the highest dimensionless frequency of interest is $(ka)_{\max}$. The estimated minimal measurements for an arbitrary surface[21] are $\tilde{M}=1.4[(ka)_{\max}+1]^2$, which is usually much less than N_0. The procedures for hybrid NAH are described as follows.

First, we take M_1 measurements of acoustic pressure on a conformal surface Γ around the source at close range. Next we use Eq. (8a) to determine the expansion coefficients $\hat{C}_j$, $j=1$ to $2J$, and Eq. (8b) to determine the optimal number of expansion functions J_{op} by minimizing reconstruction errors with respect to additional M_2 measurements on Γ, which guarantees an accurate reconstruction of the acoustic pressures on Γ. Once this is done, we employ Eq. (5) to regenerate as much acoustic pressure as needed, $\mathbf{x}_m^\Gamma\in\Gamma$, $m=1$ to M ($M\geqslant N_0$), and take them as input to (10) to reconstruct $\mathbf{p}(\mathbf{x}_S;\omega)$ and $\mathbf{v}(\mathbf{x}_S;\omega)$ at N_0 nodes ($N_0\gg M_1+M_2=\tilde{M}$).

Mathematically, we can express the solutions of Eq. (10) as

$$\begin{aligned}&\mathbf{p}(\mathbf{x}_S;\omega)=\mathbf{T}_p^\dagger(\mathbf{x}_m^\Gamma|\mathbf{x}_S;\omega)\mathbf{p}(\mathbf{x}_m^\Gamma;\omega)\\ &\text{and}\\ &\mathbf{v}(\mathbf{x}_S;\omega)=\mathbf{T}_v^\dagger(\mathbf{x}_m^\Gamma|\mathbf{x}_S;\omega)\mathbf{p}(\mathbf{x}_m^\Gamma;\omega),\end{aligned} \tag{14}$$

where $\mathbf{T}_p^\dagger=(\mathbf{T}_p^H\mathbf{T}_p)^{-1}\mathbf{T}_p^H$ and $\mathbf{T}_v^\dagger=(\mathbf{T}_v^H\mathbf{T}_v)^{-1}\mathbf{T}_v^H$ represent the pseudo inversions of $\mathbf{T}_p$ and $\mathbf{T}_v$, respectively, with a superscript H implying a conjugate transpose and $\mathbf{p}(\mathbf{x}_m^\Gamma;\omega)$ is regenerated by

$$\begin{aligned}&\mathbf{p}(\mathbf{x}_m^\Gamma;\omega)=\mathbf{A}_{M\times 2J_{\text{op}}}\tilde{\mathbf{A}}^\dagger_{2J_{\text{op}}\times\tilde{M}}\tilde{\mathbf{p}}(\mathbf{x}_{\tilde{m}}^\Gamma;\omega),\\ &\tilde{m}=1\text{ to }\tilde{M}\quad(\tilde{M}\ll N_0\leqslant M).\end{aligned} \tag{15}$$

Note that the operator $\mathbf{A}_{M\times 2J_{\text{op}}}\tilde{\mathbf{A}}^\dagger_{2J_{\text{op}}\times\tilde{M}}$ is linear and constant for a given coordinate and frequency. Hence errors contained in the input data will be transferred to the regenerated ones.

For example, if errors in $\tilde{\mathbf{p}}(\mathbf{x}_{\tilde{m}}^{\Gamma};\omega)$ are uncorrelated, so are errors in the regenerated ones $\mathbf{p}(\mathbf{x}_{m}^{\Gamma};\omega)$, provided that the coordinates of the regenerated data are different. However, the reconstruction efficiency is greatly enhanced. Equation (14) shows that the surface acoustic pressures $\mathbf{p}(\mathbf{x}_S;\omega)$ and normal velocities $\mathbf{v}(\mathbf{x}_S;\omega)$ on N_0 discrete nodes of an arbitrary source are reconstructed by $\tilde{M}$ measurements. Since $\tilde{M} \ll N_0$, the measurement complexity and time are significantly reduced.

It is important to point out that the accuracy of the regenerated field acoustic pressure is consistent with that of the measured data, provided that sufficient measurements $\tilde{M}$ are taken and that the modified HELS converges fast enough on the measurement surface. Hence there is no need to take more measurements than necessary. Needless to say, $\tilde{M}$ will increase with the dimensionless frequency ka. Reference 17 provides a formulation for estimating the number of measurements $\tilde{M}$ for a given value of ka.

IV. REGULARIZATION

In NAH applications, the input data are inaccurate and incomplete and the matrices in Eq. (12) are ill conditioned. Under this condition, a pseudo-inversion (14) will not give a convergent result and the matrices in Eq. (12) must be regularized. Over the past few decades many effective regularization methods have been developed.[34,35] However, there is no single regularization that is suitable for all inverse problems. It all depends on the problems at hand.

A basic regularization feature is to smooth the dependence of solution on the input. As an example, let us rewrite Eq. (12) as

$$\mathbf{p}^{\delta} = \mathbf{T}\mathbf{v}^{\delta}, \tag{16}$$

where $\mathbf{p}^{\delta}$ represents a set of measured acoustic pressures, $\mathbf{v}^{\delta}$ stands for acoustic quantities to be reconstructed on a source surface, and $\mathbf{T}$ is a (Hermitian) transfer matrix.

A natural way to strike a balance between the fidelity to input and that to the source field is to use Tikhonov regularization (TR), which minimizes the residuals of Eq. (16) based on some *a priori* knowledge of the source field,[36]

$$J(\mathbf{v}^{\delta,\alpha}) = \|\mathbf{T}\mathbf{v}^{\delta} - \mathbf{p}^{\delta}\|_2^2 + \alpha\|\mathbf{L}\mathbf{v}^{\delta}\|_2^2, \tag{17}$$

where the reconstructed field $\mathbf{v}^{\delta,\alpha}$ depends on the regularization parameter α and $\mathbf{L}$ is the matrix that depends on the type of regularization. Here we choose $\mathbf{L}$ to be an identify matrix. The first term on the right side of Eq. (17) indicates an ultra-rough least-squares solution for which $\alpha=0$, whereas the second term implies an ultra-smooth solution for which $\alpha\to\infty$. Thus, the choice of α allows one to decide how far to go to achieve certain smoothness.

Once the regularization format is selected, the next step is to determine the value of α. One way of specifying α that requires no prior knowledge of noise variance is to use generalized cross validation (GCV).[37,38] The basic idea of GCV is to leave a particular measured acoustic pressure out of calculations of the cost function first, and then evaluate the effectiveness of the reconstructed source field in predicting the value of the omitted data. This process is repeated for all data points and the resulting regularization parameter α can ensure a best fit of the predicted acoustic pressures to all measurements.

Using TR and GCV, we can solve Eq. (17) and express the reconstructed surface acoustic pressure $\mathbf{p}^{\delta,\alpha}$ and the surface normal velocity $\mathbf{v}^{\delta,\alpha}$ as[39]

$$\mathbf{p}^{\delta,\alpha} = \mathbf{V}_p\mathbf{F}_p^{\alpha}\boldsymbol{\Sigma}_p^{-1}\mathbf{U}_p^H\mathbf{p}^{\delta} \quad \text{and} \quad \mathbf{v}^{\delta,\alpha} = \mathbf{V}_v\mathbf{F}_v^{\alpha}\boldsymbol{\Sigma}_v^{-1}\mathbf{U}_v^H\mathbf{p}^{\delta}, \tag{18}$$

where $\mathbf{p}^{\delta,\alpha}$ and $\mathbf{v}^{\delta,\alpha}$ are α dependent, $\mathbf{U}_p$ and $\mathbf{V}_p$ are unitary matrices that contain the left and right singular vectors of $\mathbf{T}_p$, respectively, $\mathbf{U}_v$ and $\mathbf{V}_v$ are unitary matrices that contain the left and right singular vectors of $\mathbf{T}_v$, respectively, $\boldsymbol{\Sigma}_p$ and $\boldsymbol{\Sigma}_v$ are diagonal matrices containing the singular values of $\mathbf{T}_p$ and $\mathbf{T}_v$, respectively, $\mathbf{F}_p^{\alpha}$ and $\mathbf{F}_v^{\alpha}$ are the low-pass filters that are devised to eliminate the high wavenumbers in reconstructing $\mathbf{p}^{\delta,\alpha}$ and $\mathbf{v}^{\delta,\alpha}$, respectively,

$$\mathbf{F}_p^{\alpha} = \mathrm{diag}\left(\cdots, \frac{\sigma_{p,j}^2}{\alpha+\sigma_{p,j}^2}, \cdots\right)$$

and

$$\mathbf{F}_v^{\alpha} = \mathrm{diag}\left(\cdots, \frac{\sigma_{v,j}^2}{\alpha+\sigma_{v,j}^2}, \cdots\right), \tag{19}$$

where subscripts p and v in (18) and (19) imply reconstruction of acoustic pressure and normal velocity, respectively, and $\sigma_{p,j}$ and $\sigma_{v,j}$ are the jth singular values of $\mathbf{T}_p$ and $\mathbf{T}_v$, respectively.

Note that the standard TR tends to limit the growth of all wavenumber components. Since reconstruction error is primarily caused by an amplification of noise with the high wavenumber components that drop below the noise level, it is logical to suppress the high wavenumbers while leaving low wavenumbers unchanged. Therefore, we can modify the penalty function in (19) by restricting the growth of high wavenumbers and rewrite the reconstructed acoustic quantities as

$$\mathbf{p}^{\delta,\alpha} = \mathbf{V}_p\mathbf{F}_{p,M}^{\alpha}\boldsymbol{\Sigma}_p^{-1}\mathbf{U}_p^H\mathbf{p}^{\delta} \quad \text{and} \quad \mathbf{v}^{\delta,\alpha} = \mathbf{V}_v\mathbf{F}_{v,M}^{\alpha}\boldsymbol{\Sigma}_v^{-1}\mathbf{U}_v^H\mathbf{p}^{\delta}, \tag{20}$$

where $\mathbf{F}_{p,M}^{\alpha}$ and $\mathbf{F}_{v,M}^{\alpha}$ are the low-pass filters for a modified TR that are given, respectively, by[36]

$$\mathbf{F}_{p,M}^{\alpha} = \mathrm{diag}\left(\cdots, \frac{\sigma_{p,j}^2(\alpha+\sigma_{p,j}^2)^2}{\alpha^3+\sigma_{p,j}^2(\alpha+\sigma_{p,j}^2)^2}, \cdots\right)$$

and

$$\mathbf{F}_{v,M}^{\alpha} = \mathrm{diag}\left(\cdots, \frac{\sigma_{v,j}^2(\alpha+\sigma_{v,j}^2)^2}{\alpha^3+\sigma_{v,j}^2(\alpha+\sigma_{v,j}^2)^2}, \cdots\right). \tag{21}$$

The regularization parameter α for reconstructing acoustic pressure is determined by GCV[37]

$$\min\left(\frac{\|(\mathbf{I}-\mathbf{F}_{p,M}^{\alpha})\mathbf{U}_p^H\mathbf{p}^{\delta}\|_2^2}{Tr(\mathbf{I}-\mathbf{F}_{p,M}^{\alpha})^2}\right), \tag{22}$$

and that for reconstructing normal surface velocity can be written as

$$\min\left(\frac{\|(\mathbf{I}-\mathbf{F}_{v,M}^{\alpha})\mathbf{U}_v^H\mathbf{p}^{\delta}\|_2^2}{Tr(\mathbf{I}-\mathbf{F}_{v,M}^{\alpha})^2}\right), \tag{23}$$

where the symbol **I** in Eqs. (22) and (23) implies an identity matrix.

Note that one can use other regularization techniques such as standard TR or Landweber iterations[40] together with Morozov discrepancy principle (MDP)[41] or L-curve[42] to solve Eq. (12). Numerical tests demonstrate that GCV may fail to yield a value of α when it is coupled with standard TR. Although MDP can always yield a value for α, its accuracy may not be very high when it is coupled with standard TR. An optimal combination seems to be a modified TR and GCV, which is adopted in this paper.

To summarize, we reconstruct the acoustic pressure and normal velocity on an arbitrarily shaped source surface using Eq. (20) that is derived from a BEM-based integral formulation (12) regularized by modified TR and GCV and implemented through SVD. The input data to Eq. (20) are regenerated by a modified HELS formula (5), whose expansion coefficients are determined and optimized using iterations through least squares (8). Since the actual measurements in hybrid NAH are much less than the discrete nodes needed to describe the acoustic quantities on a source surface, the efficiency of reconstruction is significantly enhanced. Meanwhile, the accuracy of reconstruction is ensured by the Helmholtz integral formulations and an optimal combination of modified TR and GCV.

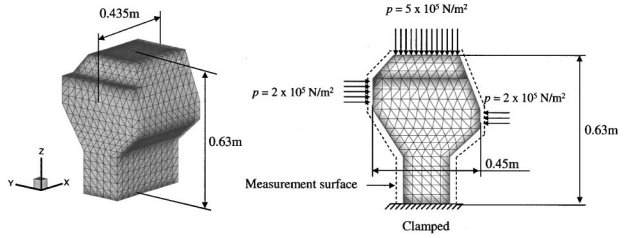


FIG. 1. Schematic of a simplified engine block subject to harmonic pressure force excitations.

Note that when hybrid NAH is used to reconstruct acoustic radiation from a sphere in a free field, we must set $D_j=0$ for there is no incoming wave. Under this condition, hybrid NAH is the same as CHELS method.[25]

V. RECONSTRUCTION OF ACOUSTIC RADIATION

We now illustrate reconstruction of acoustic radiation from an arbitrarily shaped structure vibrating in a free field via hybrid NAH. In particular, we want to demonstrate an enhancement of the efficiency and accuracy in reconstruction by using hybrid NAH as they are compared with those by using IBEM.

Figure 1 depicts the schematic of a simplified four-

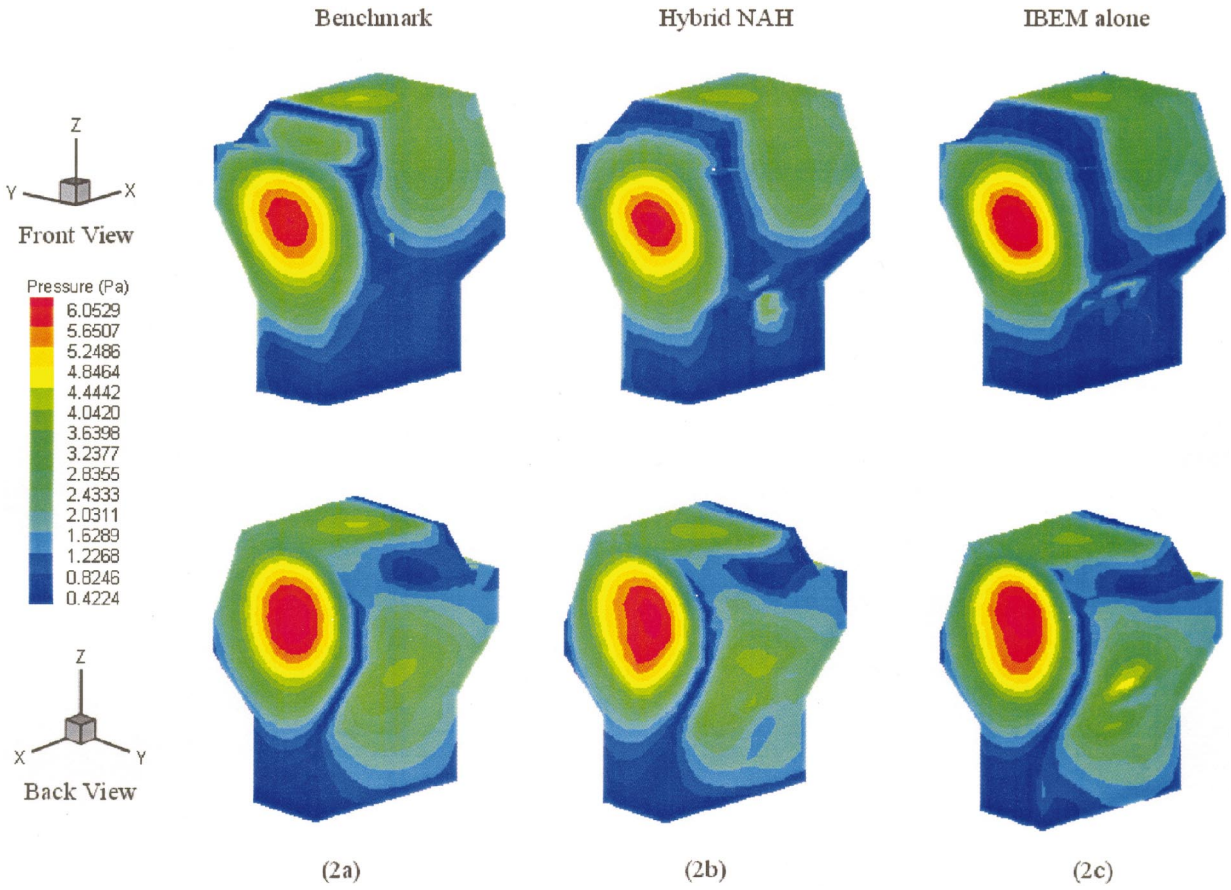


FIG. 2. Comparison of reconstructed acoustic pressures on the surface of a simplified engine block at $ka=5$: (a) Benchmark, (b) hybrid NAH with $\tilde{M}=388$, and (c) IBEM with $M=766$.

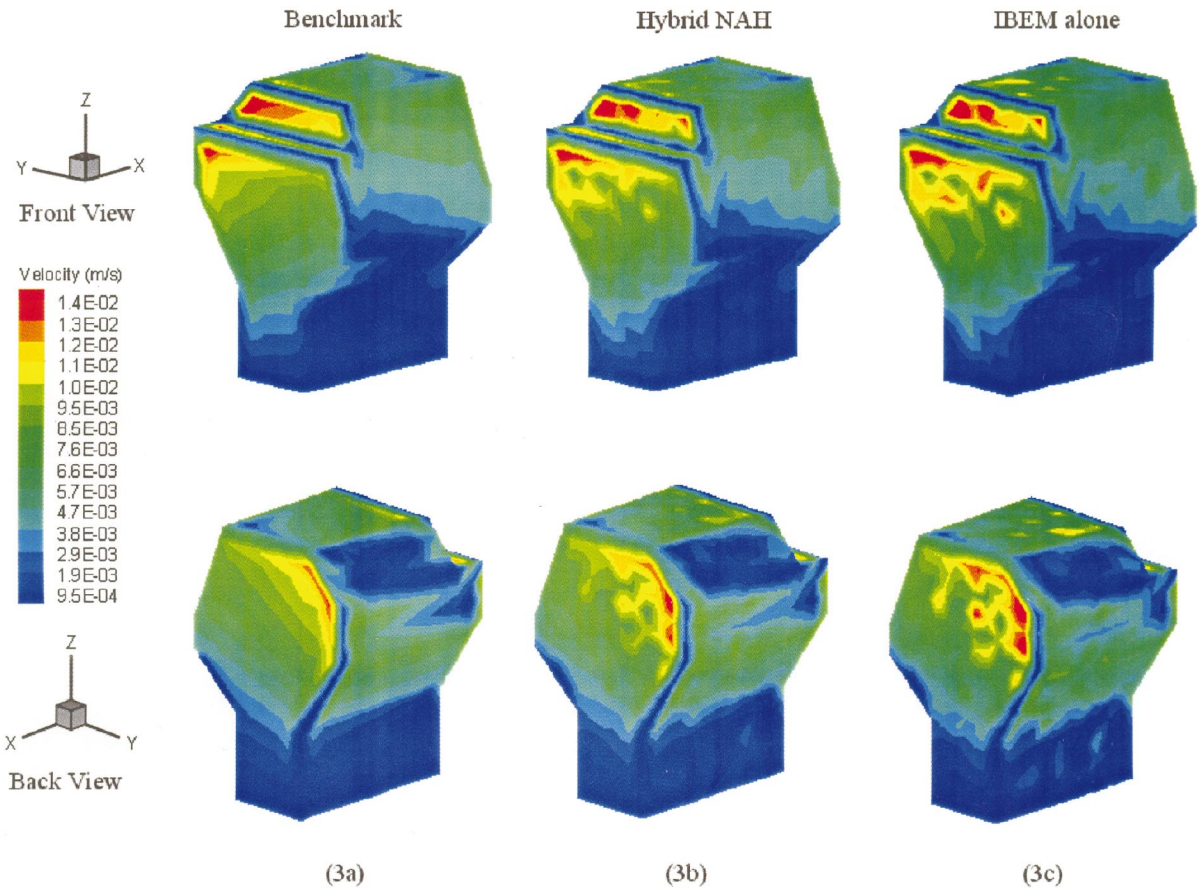


FIG. 3. Comparison of reconstructed normal velocities on the surface of a simplified engine block at $ka=5$: (a) benchmark, (b) hybrid NAH with $\tilde{M}=388$, and (c) IBEM with $M=766$.

cylinder engine block. We assume that the block is clamped at the bottom and is subject to harmonic excitations of distributed pressures on arbitrarily selected surfaces: 5×10^5 N/m^2 on the top and 2×10^5 N/m^2 on the protruding parts of the front and back surfaces, respectively. The overall dimensions of this block are 0.435 m in the x axis, 0.46 m in the y axis, and 0.63 m in the z axis direction. The normal surface velocity is obtained numerically using commercial software NASTRAN® by discretizing the engine block into 1530 six-node wedge solid elements and first-order brick solid elements with 1878 nodes. The surface acoustic pressures are calculated using Eq. (11), which is implemented numerically via BEM through discretizing the surface into 1548 linear triangular elements with 776 nodes. This surface mesh allows at least 6 nodes per wavelength for a dimensionless frequency up to $ka=5$. Note that in this study there is no additional noise added to the input data.

Once the surface acoustic quantities are specified, the field acoustic pressure is calculated using Eq. (10) over a conformal surface around the engine block at a 5-mm clearance, and taken as the input data to Eq. (9) to determine optimal expansion coefficients $\hat{C}_j$, $j=1$ to J_{op}. Next, we use Eq. (6) to regenerate as much acoustic pressure on the conformal surface as necessary and take them as the input data to Eq. (20) to reconstruct the surface acoustic quantities.

Note that if IBEM (12) is used, we need to take around 776 measurements of acoustic pressures. To examine the effectiveness of hybrid NAH, we halve the measurements to 388 and take them as input to reconstruct surface acoustic pressures. These results are compared with those reconstructed by IBEM using 776 input data, and the benchmark values generated by Eq. (10). Similarly, the normal surface velocities reconstructed by hybrid NAH using 388 input data are compared with those reconstructed by IBEM using 776 input data, and the benchmark values produced by FEM codes. Tests are conducted for dimensionless frequencies $ka=1$ to 5. For brevity, we display the surface acoustic pressure and normal surface velocity distributions at $ka=5$, for which the optimal number of expansion functions in hybrid NAH is $J_{\text{op}}=39$.

Figures 2 and 3 depict comparisons of the surface acoustic pressures and normal surface velocities reconstructed by hybrid NAH using 388 input data and those by IBEM using 776 input data versus the benchmark values produced by BEM and FEM, respectively. These results illustrate that hybrid NAH can yield similar accuracy and resolution in reconstruction as IBEM does, but only needs half of the measurements to get the job done.

Figure 4 displays the first five orthonormal eigenfunctions or modes of the normal velocity on the engine block surface at $ka=5$ obtained by hybrid NAH. Each mode $V_v^{(i)}$ represents the singular vector of the unitary matrix $\mathbf{V}_v^H$ associated with the transfer matrix $\mathbf{T}_v$ (here $i=1$ to 9 implies the ith singular value of $\mathbf{T}_v$). These modes decompose the nor-

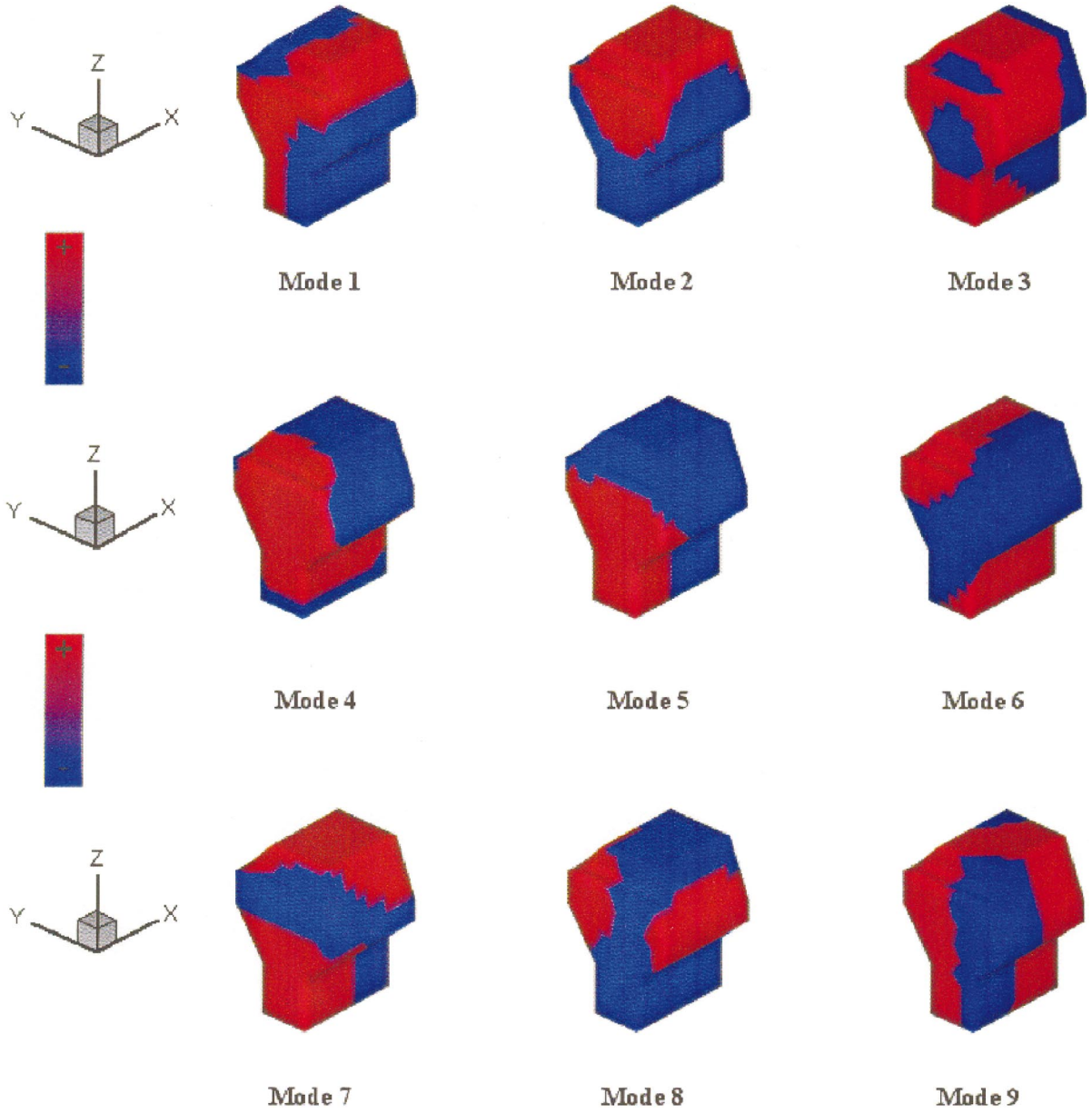


FIG. 4. The first nine modes of the normal velocity on the engine block surface at $ka=5$ (red: positive; blue: negative).

mal surface velocity vector of a simplified engine block. By taking discrete Fourier transform of the normal surface velocity, we can separate the supersonic components and identify the modes that are responsible for far-field acoustic radiation from a vibrating structure.[43]

For completeness, we depict the surface acoustic quantities reconstructed by IBEM using the same 388 input data as hybrid NAH does at $ka=5$. Since the discrete nodes are twice as many as measurements, Eq. (12) is severely underdetermined. Figure 5 shows the SVD spectra of $\mathbf{T}_p$ and $\mathbf{T}_v$ under 776 and 386 input data, respectively. Note that the surface velocities usually contain more near-field effects than surface pressures do, and $\mathbf{T}_v$ is more ill conditioned than $\mathbf{T}_p$ is. So when $\mathbf{T}_v$ is underdetermined, we can still find the right regularization parameter and filter out small singular values to get a satisfactory reconstruction. This is not the case for the surface acoustic pressure because $\mathbf{T}_p$ is weakly ill conditioned. As a result, taking fewer measurements (having an

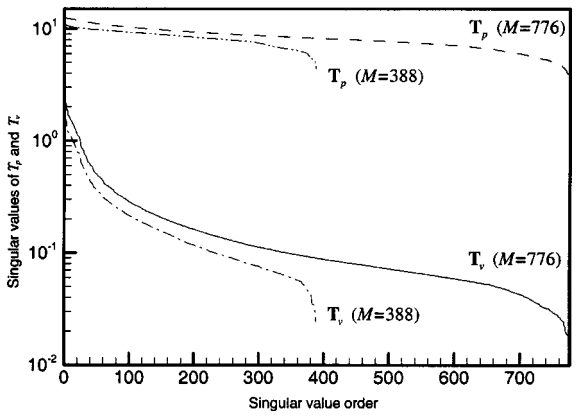


FIG. 5. SVD spectra of the transfer matrices $\mathbf{T}_p$ and $\mathbf{T}_v$. – – –: $\mathbf{T}_p$ with 776 input data; –·–·–: $\mathbf{T}_p$ with 388 input data; ———: $\mathbf{T}_v$ with 776 input data; –·–: $\mathbf{T}_v$ with 388 input data.

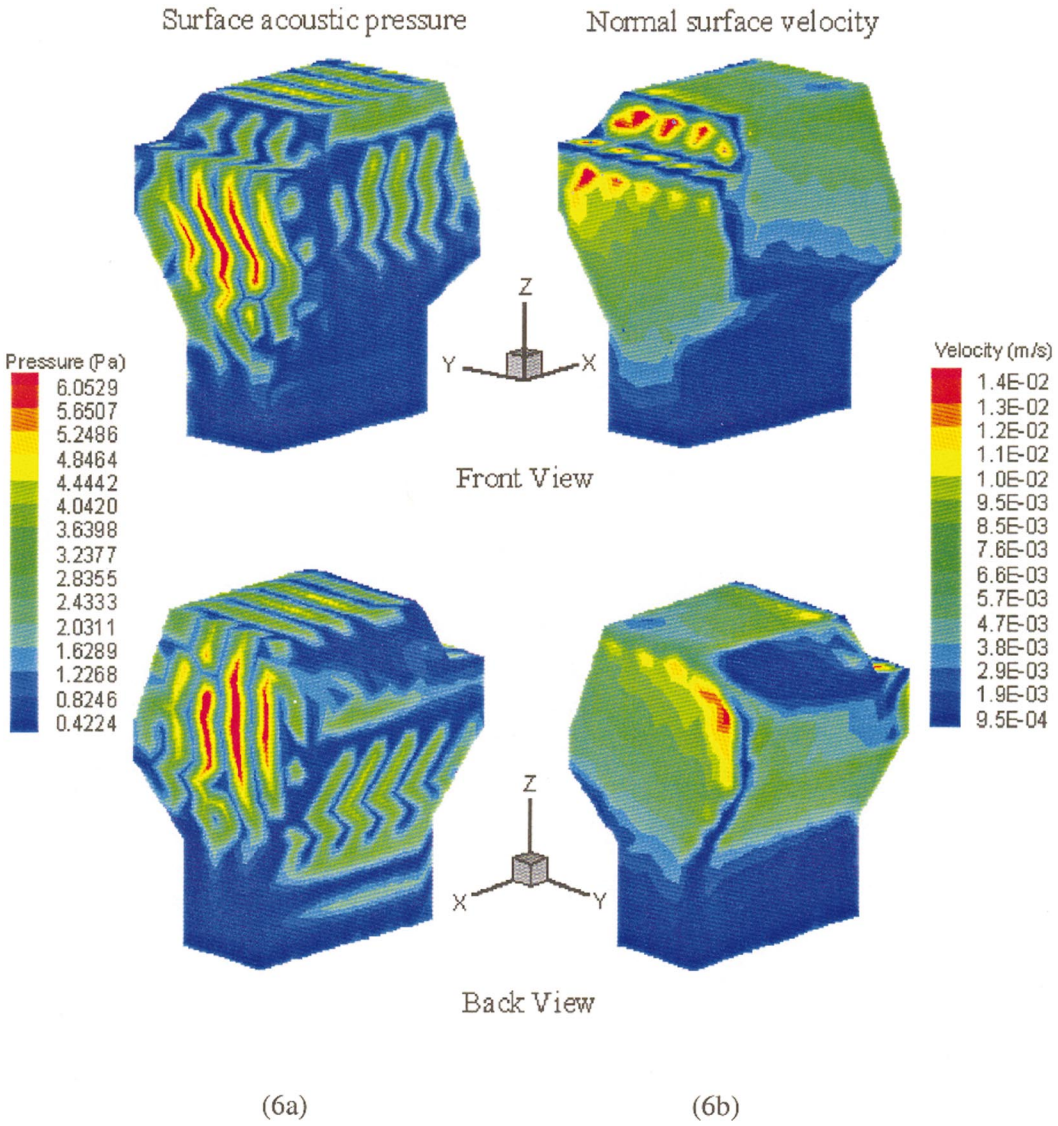


FIG. 6. The surface acoustic quantities reconstructed by using IBEM with $\tilde{M}=388$ at $ka=5$: (a) surface acoustic pressures and (b) normal surface velocities.

underdetermined system) means not getting enough information in the input. Under this condition, there is no way of reconstructing the surface acoustic pressure correctly even if regularization is employed. Oftentimes, the reconstructed image is distorted (see Fig. 6).

Figure 7 shows sums of relative error norms in reconstructing surface acoustic pressures $\|\mathbf{p}_{\text{bench}}-\mathbf{p}^{\delta,\alpha}\|_2^2/\|\mathbf{p}_{\text{bench}}\|_2^2 \times 100\%$ with respect to all 776 nodes by hybrid NAH using 388 input and by IBEM using 776 and 388 input data, respectively, for $ka=1$ to 5. Results indicate that when 388 measurements are fed to IBEM, errors in reconstructing surface acoustic pressure are greatly increased. A similar trend is observed in calculating sums of relative error norms for reconstructing normal surface velocities $\|\mathbf{v}_{\text{bench}}-\mathbf{v}^{\delta,\alpha}\|_2^2/\|\mathbf{v}_{\text{bench}}\|_2^2 \times 100\%$ (see Fig. 8). Note that reconstruction errors in hybrid NAH using 388 data are higher than those in IBEM using 776 input data. This is a penalty for hybrid NAH to reduce the measurements. Considering the complexities involved in taking a large number of measurements in practice, however, such a penalty is worth taking.

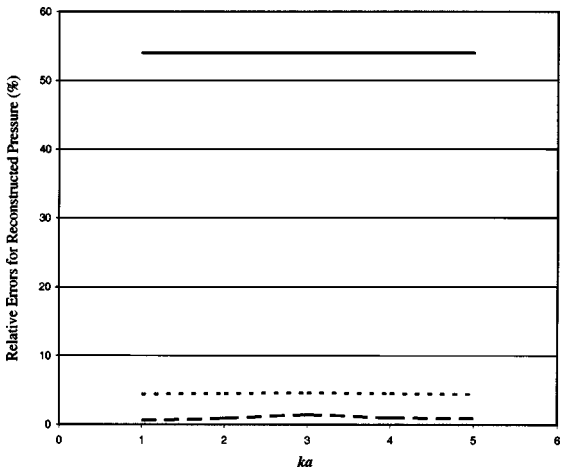


FIG. 7. Relative error norms for the reconstructed acoustic pressures on the surface of a simplified engine block. ———: IBEM with $\tilde{M}=388$; ···: hybrid NAH with $\tilde{M}=388$; and – – –: IBEM with $M=766$.

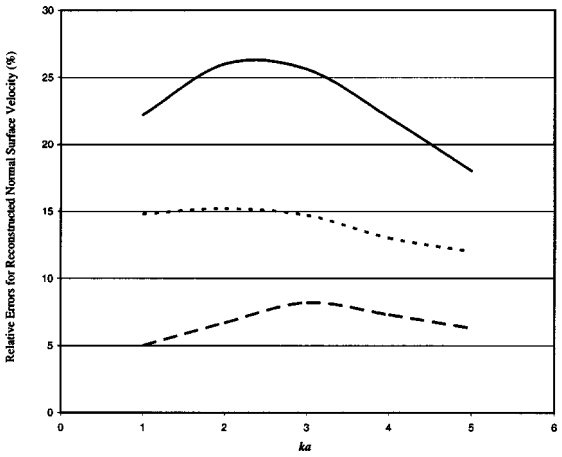


FIG. 8. Relative error norms for the reconstructed normal velocities on the surface of a simplified engine block. ———: IBEM with $\tilde{M}=388$; ···: hybrid NAH with $\tilde{M}=388$; and – – –: IBEM with $M=766$.

VI. CONCLUSIONS

Hybrid NAH is developed for reconstructing acoustic radiation from arbitrary vibrating structures. The basis of this hybrid NAH is a modified HELS that expresses the radiated acoustic pressure in terms of the outgoing and incoming spherical wave functions. This modified HELS is combined with the Helmholtz integral formulations implemented by BEM, which are regularized by a modified TR and GCV and implemented via SVD. These hybrid formulas are applicable for an arbitrary object in a confined and free fields. In applying this hybrid NAH to a confined field, measurements should be taken on two conformal surfaces around a source in order to discern the directions of wave propagations, when the boundary surfaces are highly reflective. For a source in a free field, it is sufficient to take measurements on a single conformal surface at close range.

Numerical results depict that this hybrid NAH can yield a satisfactory reconstruction of surface acoustic quantities with fewer measurements than IBEM does. When the same input data to hybrid NAH are fed to IBEM, the system of equations is severely underdetermined. Although in general we can get bounded solutions via regularization for an underdetermined system, the accuracy of solutions depends on the ill posedness of a problem. The more ill posed a problem is, the more likely it is to get the regularization parameter to obtain a satisfactory reconstruction for an underdetermined system. The less ill posed a problem is, the less likely it is to get a correct reconstruction for an underdetermined system. This is because for a weakly ill-posed problem, taking fewer measurements means losing vital information in the input. Since we do not know *a priori* how ill posed a problem is and since lacking sufficient information is a major cause of the ill posedness in reconstruction, it is a good idea to take enough measurements in NAH practice.

ACKNOWLEDGMENTS

The author is indebted to Dr. Xiang Zhao for his assistance in preparing numerical results presented in this paper. Special thanks also go to Tatiana Semenova for providing invaluable comments during the preparation. This work was supported by NSF Grant No. ITR/ACS-0081270.

[1] E. G. Williams and J. D. Maynard, ''Holographic imaging without the wavelength resolution limit,'' Phys. Rev. Lett. **45**, 554–557 (1980).
[2] E. G. Williams, J. D. Maynard, and E. Skudrzky, ''Sound source reconstructions using a microphone array,'' J. Acoust. Soc. Am. **68**, 340–344 (1980).
[3] E. G. Williams, H. D. Dardy, and R. G. Fink, ''Nearfield acoustical holography using an underwater automated scanner,'' J. Acoust. Soc. Am. **78**, 789–798 (1985).
[4] E. G. Williams, ''The nearfield acoustic holography (NAH) experimental method applied to vibration and radiation in light and heavy fluids,'' Comput. Struct. **65**, 323–335 (1997).
[5] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, San Diego, 1999).
[6] E. G. Williams, H. D. Dardy, and K. B. Washburn, ''Generalized nearfield acoustical holography for cylindrical geometry: Theory and experiment,'' J. Acoust. Soc. Am. **81**, 389–407 (1987).
[7] J. D. Maynard, E. G. Williams, and Y. Lee, ''Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH,'' J. Acoust. Soc. Am. **78**, 1395–1413 (1985).
[8] W. A. Veronesi and J. D. Maynard, ''Nearfield acoustic holography (NAH): II. Holographic reconstruction algorithms and computer implementation,'' J. Acoust. Soc. Am. **81**, 1307–1322 (1987).
[9] B. K. Gardner and R. J. Bernhard, ''A noise source identification technique using an inverse Helmholtz integral equation method,'' ASME J. Vib., Acoust., Stress, Reliab. Des. **110**, 84–90 (1988).
[10] W. A. Veronesi and J. D. Maynard, ''Digital holographic reconstruction of sources with arbitrarily shaped surfaces,'' J. Acoust. Soc. Am. **85**, 588–598 (1989).
[11] G. V. Borgiotti, A. Sarkissian, E. G. Williams, and L. Schuetz, ''Conformal generalized near-field acoustic holography for axisymmetric geometries,'' J. Acoust. Soc. Am. **88**, 199–209 (1990).
[12] G.-T. Kim and B.-T. Lee, ''3-D sound source reconstruction and field reproduction using the Helmholtz integral equation,'' J. Sound Vib. **136**, 245–261 (1990).
[13] M. R. Bai, ''Application of BEM (boundary element method)-based acoustic holography to radiation analysis of sound sources with arbitrarily shaped geometries,'' J. Acoust. Soc. Am. **92**, 533–549 (1992).
[14] B.-K. Kim and J.-G. Ih, ''On the reconstruction of vibro-acoustic field over the surface enclosing an interior space using the boundary element method,'' J. Acoust. Soc. Am. **100**, 3003–3016 (1996).
[15] Z. Zhang, N. Vlahopoulos, S. T. Raveendra, T. Allen, and K. Y. Zhang, ''A computational acoustic field reconstruction process based on an indirect boundary element formulation,'' J. Acoust. Soc. Am. **108**, 2167–2178 (2000).
[16] Z. Wang and S. F. Wu, ''Helmholtz equation-least-squares method for reconstructing the acoustic pressure field,'' J. Acoust. Soc. Am. **102**, 2020–2032 (1997).
[17] S. F. Wu, ''On reconstruction of acoustic pressure fields using the Helmholtz equation least squares method,'' J. Acoust. Soc. Am. **107**, 2511–2522 (2000).
[18] H. A. Schenck, ''Improved integral formulation for acoustic radiation problems,'' J. Acoust. Soc. Am. **44**, 41–58 (1968).
[19] D. Colton and R. Kress, *Integral Equation Methods in Scattering Theory* (Wiley, New York, 1983).
[20] N. E. Rayess and S. F. Wu, ''Experimental validations of the HELS method for reconstructing acoustic radiation from a complex vibrating structure,'' J. Acoust. Soc. Am. **107**, 2955–2964 (2000).
[21] S. F. Wu, N. Rayess, and X. Zhao, ''Visualization of acoustic radiation from a vibrating bowling ball,'' J. Acoust. Soc. Am. **109**, 2771–2779 (2001).
[22] S. F. Wu and J. Yu, ''Reconstructing interior acoustic pressure field via Helmholtz equation least-squares method,'' J. Acoust. Soc. Am. **104**, 2054–2060 (1998).
[23] M. Moondra and S. F. Wu, ''Visualization of vehicle interior noise using HELS based NAH,'' (submitted to) Noise Control Engi. J.
[24] E. G. Williams, ''Continuation of acoustic near-fields,'' J. Acoust. Soc. Am. **113**, 1273–1281 (2003).
[25] S. F. Wu and X. Zhao, ''Combined Helmholtz equation least squares (CHELS) method for reconstructing acoustic radiation,'' J. Acoust. Soc. Am. **112**, 179–188 (2002).
[26] B.-D. Lim and S. F. Wu, ''Determination of the optimal number of expan-

sion terms in the HELS method,'' J. Acoust. Soc. Am. **108**, 2505 (2000).
[27] I. N. Vekua, ''On completeness of a system of metaharmonic functions,'' Dokl. Akad. Nauk SSSR **90**(5), 715–718 (1953) (in Russian).
[28] R. F. Millar, ''The Rayleigh hypothesis and a related least-squares solution to scattering problems for periodic surfaces and other scatterers,'' Radio Sci. **8**(8), 785–796 (1973).
[29] R. F. Millar, ''On the completeness of sets of solutions to the Helmholtz equation,'' IMA J. Appl. Math. **30**, 27–37 (1983).
[30] V. Isakov and S. F. Wu, ''On theory and applications of the HELS method in inverse acoustics,'' Inverse Probl. **18**, 1147–1159 (2002).
[31] N. E. Rayess, ''Development of acoustic holography using the Helmholtz Equation-Least Squares (HELS) Method,'' Ph.D. dissertation, Department of Mechanical Engineering, Wayne State University, May, 2001.
[32] M. Moondra and S. F. Wu, ''Visualization of vehicle interior noise using HELS based NAH,'' (CD-ROM) Proceedings of the INTER-NOISE 2002, August, 2002.
[33] A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981), Chap. 4, pp. 165–194.
[34] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems* (Wiley, New York, 1977), Chap. 2, pp. 71–73.
[35] S. Twomey, ''On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature,'' J. Assoc. Comput. Mach. **10**, 97–101 (1962).
[36] E. G. Williams, ''Regularization methods for nearfield acoustical holography,'' J. Acoust. Soc. Am. **110**, 1976–1988 (2001).
[37] G. H. Golub, M. Heath, and G. Wahba, ''Generalized cross-validation as a method for choosing a good ridge parameter,'' Technometrics **21**, 215–223 (1979).
[38] S. H. Yoon and P. A. Nelson, ''Estimation of acoustic source strength by inverse methods: Part II, Experimental investigation of methods for choosing regularization parameters,'' J. Sound Vib. **233**, 669–705 (2000).
[39] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems* (SIAM, Philadelphia, PA, 1998).
[40] L. Landweber, ''An iteration formula for Fredholm integral equations of the first kind,'' Am. J. Math. **73**, 615–624 (1951).
[41] V. A. Morozov, ''The error principle in the solution of operational equations by the regularization method,'' USSR Comput. Math. Math. Phys. **8**, 63–87 (1968).
[42] P. C. Hansen and D. P. O'Leary, ''The use of the L-curve in the regularisation of discrete ill-posed problems,'' SIAM J. Sci. Comput. (USA) **14**, 1487–1503 (1993).
[43] X. Zhao and S. F. Wu, ''Reconstruction of the vibro-acoustic fields using hybrid nearfield acoustical holography,'' submitted to the J. Sound Vib. (2003).

Sound transmission characteristics of Tee-junctions and the associated length corrections

S. K. Tang[a)]
Department of Building Services Engineering, The Hong Kong Polytechnic University, Hong Kong, China

(Received 30 March 2003; revised 22 August 2003; accepted 17 October 2003)

The sound transmission characteristics of a Tee-junction formed by a sidebranch and an infinitely long duct are investigated numerically using the finite element method. The associated corrections of the branch length and the upstream duct length are also discussed in detail. The types of branch resonance that result in strong or weak sound transmission across the junction are determined and their effects on the length corrections examined. Results suggest that the type of sidebranch, the branch width, the branch length, and the order and the form of the resonance affect more significantly the length corrections of the duct section. The excitation of nonplanar higher branch modes gives rise to rapid increase in the duct length corrections and also results in lower sound transmission. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1631830]

PACS numbers: 43.50.Gf, 43.20.Mv [DKW] Pages: 218–227

I. INTRODUCTION

Sidebranches are very commonly found in modern buildings with centralized air conditioning systems. They help the distribution of treated air from the air handling units into the occupied zones inside a building. It has been shown that a sidebranch offers a change in the acoustic impedance along the main duct and thus results in sound transmission loss.[1] A formula for the prediction of this sound transmission loss can also be found in textbooks (for instance, Ref. 2). However, this formula is accurate only at very low frequency. Also, such a change in acoustic impedance alters the resonant characteristics along the main duct section upstream of the sidebranch. A length correction for this section is expected.

The issue of the end correction for plane wave motion for duct openings with and without flanges is well known.[1] This correction affects the resonance frequencies of the duct. The larger the opening, the greater will be this correction. However, the corresponding correction due to diffraction and/or scattering at a discontinuity internal to a duct is rarely discussed. The theoretical deduction of Miles[3] using the conformal mapping technique works only at very low frequency. Bruggeman[4] assumed plane waves inside the sidebranches and the main duct except in a limited region inside the junction. This assumption is valid only at low frequency. Neither the results of Bruggeman[4] nor the more recent results from Redmore and Mulholland[5] enable an estimation of such a correction. Knowledge of this correction in a reasonable frequency range will enable a better prediction of the sound transmission loss and provide clues for the physical explanation of acoustical properties of multiconnected duct elements along a duct.

Recently, the experimental results of the author on sound transmission across a single sidebranch revealed the presence of two main branch impedance types.[6] The first one is a high-pass filter type. This corresponds to the case of an infinitely long branch or zero branch length opening. The second one involves the longitudinal resonance inside the branch and is of the band-stop filter type. It has also been shown that a specific correction in the separation between two identical sidebranches is required to explain the measured sound power transmission loss. However, an exact explanation has not been sought, as the knowledge of the length correction due to diffraction at the branch entrances is currently unknown.

In the present study, the sound field and the acoustic length correction due to the presence of a Tee-junction in a duct are investigated using the method of finite elements. This method can allow for a much higher degree of nonuniformity in pressure/velocity distribution inside the sidebranches than that of Bruggeman.[4] The frequency concerned is up to the first eigenfrequency of the duct, so that plane wave propagation inside the duct at distances sufficiently far away from the branch is guaranteed. It is hoped that the present results can provide a stepping stone for investigation of the acoustical properties of duct element systems with more complicated connections.

II. COMPUTATIONAL MODEL AND THEORY

Figure 1 illustrates the schematic of the sidebranch and duct system (Tee-junction) and the nomenclature adopted. All length scales are normalized by the duct width d. The branch length l and the branch width w are allowed to vary relative to d. As only plane incident waves will be considered in the present study, the sound field dealt with here is uniform in the direction normal to the x-y plane. The finite element computations for solving the standard wave equation

$$\nabla^2 p + k^2 p = 0, \tag{1}$$

where p and k denote the acoustic pressure and the wave number of the sound, respectively, are implemented using the software MATLAB. The computation domain starts at $x/d = -5$ and ends at $x/d = w/d + 6$. These locations are far enough for all evanescent waves scattered and/or diffracted by the sidebranch to decay completely. At $x/d = -5$, the

[a)]Electronic mail: besktang@polyu.edu.hk

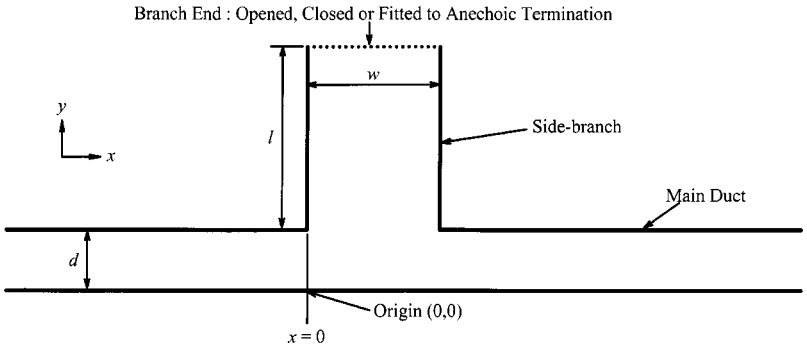


FIG. 1. Schematic of the Tee-junction and nomenclature adopted.

sound field is made up of a forward and a backward propagating plane wave, while that at $x/d=w/d+6$ consists of a forward propagating plane wave only. The acoustic pressures in the foregoing discussion are normalized by the incident wave magnitude. As in Ref. 7, the corresponding boundary conditions are

$$\frac{\partial p}{\partial n}+ikp=\begin{cases}2ik & \text{at } x/d=-5,\\ 0 & \text{at } x/d=w/d+6,\end{cases} \tag{2}$$

where n represents the outward normal direction of a boundary. The boundary conditions on boundaries other than the branch exit follow that of a rigid wall of vanishing normal acoustic particle velocity. The sound frequency is kept below the first higher-mode eigenfrequency of the main duct.

The boundary condition at the branch exit substantially affects the type of filtering action the sidebranch offers to the wave propagating along the main duct. In the present study, three types of branch endings will be considered. The first type refers to an infinitely long branch, which does not produce sound reflection back into the main duct. The impedance-matched anechoic termination of Tang and Lau[7] is adopted at $y/d=6$, $0<x<w$, instead of the plane wave condition in order to attenuate any higher branch modes and evanescent waves that may have been created in the Tee-junction. However, one should note that the branch impedance in this case does not depend on the branch length l. Such a branch produces a high-pass filtering action until the cuton of the first higher mode in the branch or in the main duct. The second type is an opened branch. For simplicity, the condition of $p=0$ at $y=l+d$, $0<x<w$ is adopted. In this case, l includes the open-end correction due to exit radiation impedance.[1] A high-pass filtering action is created before resonance occurs. The last type is the closed branch.[8] The boundary condition at $y=l+d$, $0<x<w$ is $\partial p/\partial y=0$. This branch produces low-pass filtering action up to the first longitudinal branch resonance frequency. These branch end conditions cover those categorized by Tang and Li.[6]

Although one expects the sound field to be nonplanar in the vicinity of the Tee-junction, the waves far away from the junction are still planar. The acoustic length correction for the upstream duct section due to the Tee-junction, ϵ, is determined by the wave number or frequency at which the input impedance Z_{in} to the system has vanishing reactance and its magnitude reaches a local minimum in the frequency variation.[1] The input impedance can be estimated once the acoustic pressure field at the system inlet ($x/d=-5$) and its gradient are obtained:

$$\frac{Z_{\text{in}}}{\rho cS}=-i\,\frac{kp_{x/d=-5}}{\partial p/\partial x|_{x/d=-5}}, \tag{3}$$

where ρ is the ambient air density, c the speed of sound, and S the cross-sectional area of the main duct, which is not important in the present study. The wave number k_j at which the jth order resonance occurs can then be found. They are related to the acoustical length correction ϵ_j by

$$k_j=\frac{j\pi}{L+\epsilon_j}\Rightarrow\epsilon_j=\frac{j\pi}{k_j}-L, \tag{4}$$

where L is the distance from the branch at which the duct system is being driven and is equal to $5d$ in the present study. L does not affect the length corrections.[1] The frequencies of longitudinal branch resonances determined by the computation are used to estimate the length corrections for the branches in a similar manner with L replaced by l. Since the computational results for acoustic pressure are normalized by the incident sound wave magnitude, the squared strength of the transmitted sound wave, $|p_{x/d=w/d+6}|^2$, equals the sound power transmission coefficient, which is denoted by τ in the foregoing discussion.

The maximum branch width included here is $1.8d$. This is already outside the practical dimension, which is usually $w/d\leqslant 1$. However, the range of $1<w/d\leqslant 1.8$ is included for the sake of completeness. The branch length investigated is up to $5d$.

III. RESULTS AND DISCUSSION

A. Infinitely long branches

The acoustical properties of infinitely long branches do not depend on the branch length l. However, it would be better to have a length not less than $6d$ so that the computational requirement of the numerical anechoic termination is less demanding. No standing wave is set up in the branches throughout the numerical experiment.

Figure 2 shows the variation of sound power transmission along the main duct with wave number (frequency) before the first eigenfrequency of the duct. The importance of the branch width w is manifested. For $w/d<1$, the branch is a high-pass filter, and the sound transmission becomes more and more effective as kd approaches π. For $w/d>1$, the first higher branch mode is excited before that of the main duct. Again τ increases with frequency, but this time with a faster rate than that for $w/d<1$, and reaches a local maximum at or very close to the eigenfrequency of the branch. After that, τ

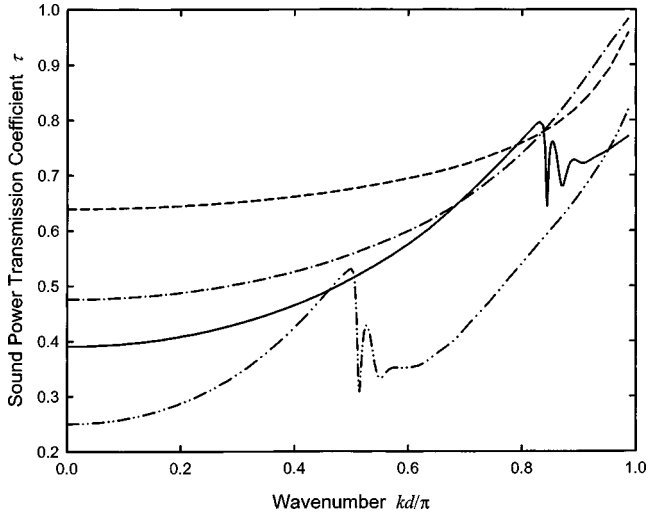


FIG. 2. Sound power transmission across an infinitely long sidebranch along the main duct. ——: $w/d=0.5$; —·—·—·—: $w/d=0.9$; ——: $w/d=1.2$; –··–: $w/d=2$.

fluctuates chaotically before it resumes its increasing trend as kd approaches the upper limit of the present study. This observation appears consistent with the experimental observation of Redmore and Mulholland,[5] which shows that the sound power transmission loss along the main duct around the first higher branch mode eigenfrequency appears at frequency slightly higher than this frequency. This will be discussed further later. For $w/d>1$, the maximum τ before the cuton of the first branch asymmetric higher mode decreases as w increases. At very low frequency, $kd\rightarrow 0$, and one can observe that

$$\tau=\left(\frac{d}{d+0.5w}\right)^2, \tag{5}$$

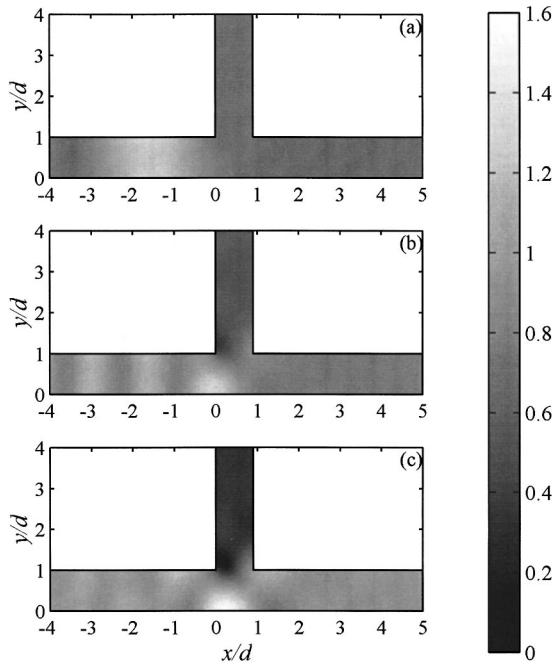


FIG. 3. Sound field around Tee-junction of an infinitely long sidebranch of $w/d=0.9$. $kd=$ (a) 0.254π, (b) 0.729π, and (c) 0.923π.

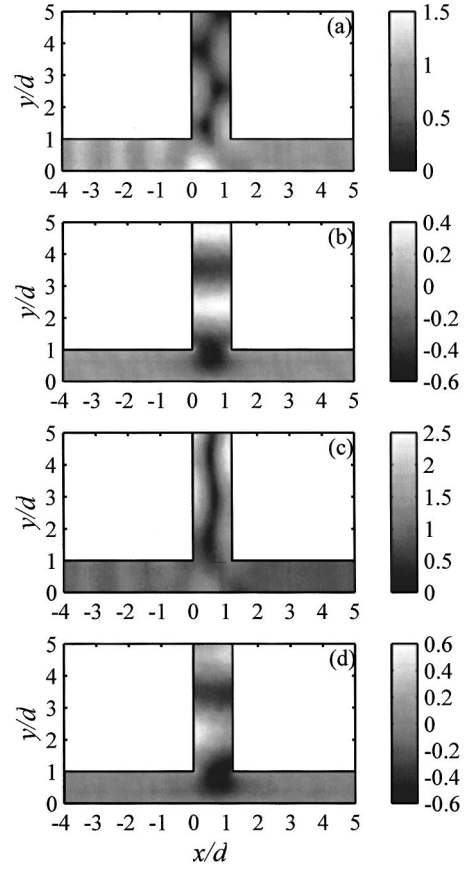


FIG. 4. Sound field around Tee-junction of an infinitely long sidebranch of $w/d=1.2$. (a) $kd=0.833\pi$, pressure magnitude; (b) $kd=0.833\pi$, upward particle velocity; (c) $kd=0.844\pi$, pressure magnitude; (d) $kd=0.844\pi$, upward particle velocity.

which agrees with the prediction of the low-frequency plane wave theory[1] regardless the ratio w/d.

Figure 3 illustrates some sound fields around the Tee-junction for $w/d=0.9$. At low frequency, the sound field close to the entry of the Tee-junction is planar, while a strong standing wave is observed in the upstream duct, suggesting that a strong reflection exists [Fig. 3(a)]. As the frequency increases, greater sound energy is accumulated inside the junction [Fig. 3(b)] due to the higher diffraction effectiveness at higher frequency.[9] This produces stronger excitation in the plane wave propagation into the downstream duct and thus reduces the sound power transmission loss. At frequencies close to the first higher eigenfrequency of the duct, some nonplanar sound waves can be found within and at some distances downstream of the junction because of the acoustic scattering and diffraction as expected [Fig. 3(c)]. These evanescent waves decay quickly inside the duct and the branch.

Sound fields for $w/d>1$ are more complicated, although those at frequencies before the first higher eigenfrequency of the branch resemble very much those shown in Fig. 3. The characteristics of the sound field development for all w larger than d are essentially the same and thus only the case of $w/d=1.2$ will be illustrated. As the sound frequency approaches the first eigenfrequency of the branch with $w/d=1.2$ ($kd=\pi/1.2\sim 2.62$), one can observe that the sound energy gradually fills up the inlet of the branch [Fig. 4(a)] and the distribution of acoustic velocity (upward) across this

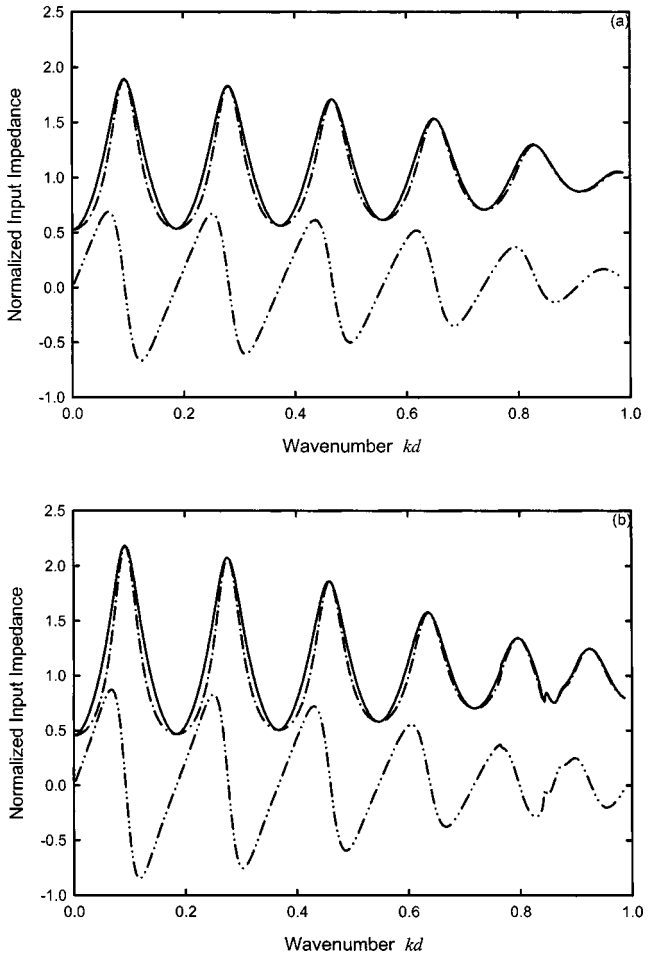


FIG. 5. Frequency variation of input impedance with an infinitely long sidebranch. $w/d=$ (a) 0.9; (b) 1.2. ——: Impedance magnitude; —·—·—·—: resistance; —··—: reactance.

inlet at frequency very close to the mentioned branch eigenfrequency is symmetrical about the branch axis [Fig. 4(b)]. Therefore, the first higher branch mode, which is asymmetric, is not properly excited even when the frequency reaches the mode eigenfrequency. At higher frequency, the wavelength of the sound becomes smaller than $2w$. Strong resonance of the higher branch mode can then be observed [Fig. 4(c)]. This strong resonance takes away a considerable amount of acoustic energy, resulting in a drop of the sound power transmission (Fig. 2). The corresponding upward particle velocity field, shown in Fig. 4(d), illustrates the asymmetric excitation at the branch inlet around this critical frequency.

Figures 5(a) and 5(b) illustrate the frequency variations of the normalized input impedance $Z_{\text{in}}/\rho cS$ for $w/d=0.9$ and 1.2, respectively. The local minima of the impedance magnitude occur at frequencies of vanishing reactance for both w. At very low frequency, that is, $kd\rightarrow 0$, the present results suggest that Z_{in} is real and conforms to the low-frequency theory[1] prediction that

$$Z_{\text{in}}=\rho cS\frac{d}{w+d}. \tag{6}$$

Figure 5 also shows that the reactance tends to converge toward zero and the resistance toward unity as the frequency

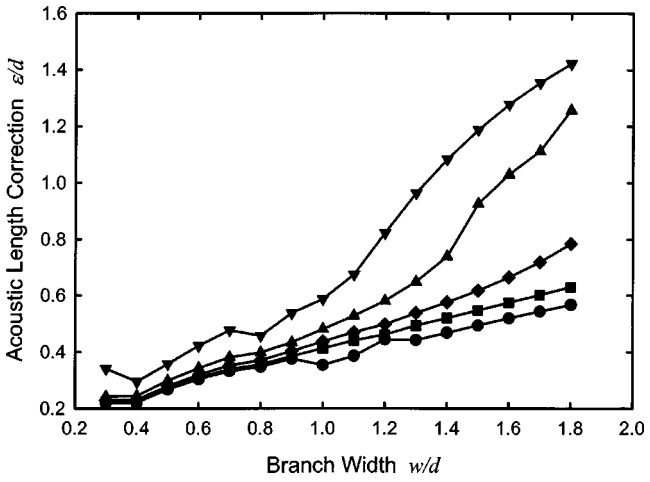


FIG. 6. Acoustic length corrections in the presence of an infinitely long sidebranch. ●: first order; ■: second order; ◆: third order; ▲: fourth order; ▼: fifth order.

increases toward the eigenfrequency of the first asymmetric higher duct mode. These features of Z_{in} are observed for other w/d. The very high-power transmission at frequency close to the first duct mode eigenfrequency, as illustrated in Fig. 2, is accompanied by a large reduction in the reflected wave energy as the resistance to wave propagation becomes weak. The situation becomes similar to that of an infinitely long duct without the sidebranch and thus $Z_{\text{in}}\rightarrow\rho cS$. This is consistent with the computed results even with the presence of evanescent waves in the vicinity of the Tee-junction. The propagation of higher branch modes does not offset the general trend of Z_{in} variation, except for the minor smooth but relatively rapid fluctuations close to the first branch mode eigenfrequency [Fig. 5(b)].

Figure 6 shows the variation of the acoustic length corrections for the first five order resonances of the main duct with wave number. Unlike the case for the open-ended tube, these corrections increase with frequency as diffraction and wave scattering are in general more efficient at higher frequency.[9] They also increase with w in general, although there are occasional irregularities in the increasing trend. Such irregularities are probably due to the very frequency and geometry dependent diffraction, which leads to nonuniform distribution of acoustic pressure and air particle velocities near the branch entrance. For small w, the first four order length corrections are very similar. The difference increases with w, and this increase accelerates when the higher branch mode starts to propagate. The fifth order resonance at $w/d=1.2$ occurs after this higher-mode cuton, while at $w/d=1.5$, the fourth one occurs. Those lower-order resonances occur before such a cuton in the present range of w/d. The propagation of higher branch modes enhances the spreading of acoustical energy across the Tee-junction. It is also noted that the first three order length corrections and those of the other two orders before the higher-mode cuton increase slowly with w, and the ratio ϵ/w actually drops as w/d increases (not shown here).

B. Opened branches

Standing waves can be found in the branches in this case due to open-end reflection. Two types of resonance, namely,

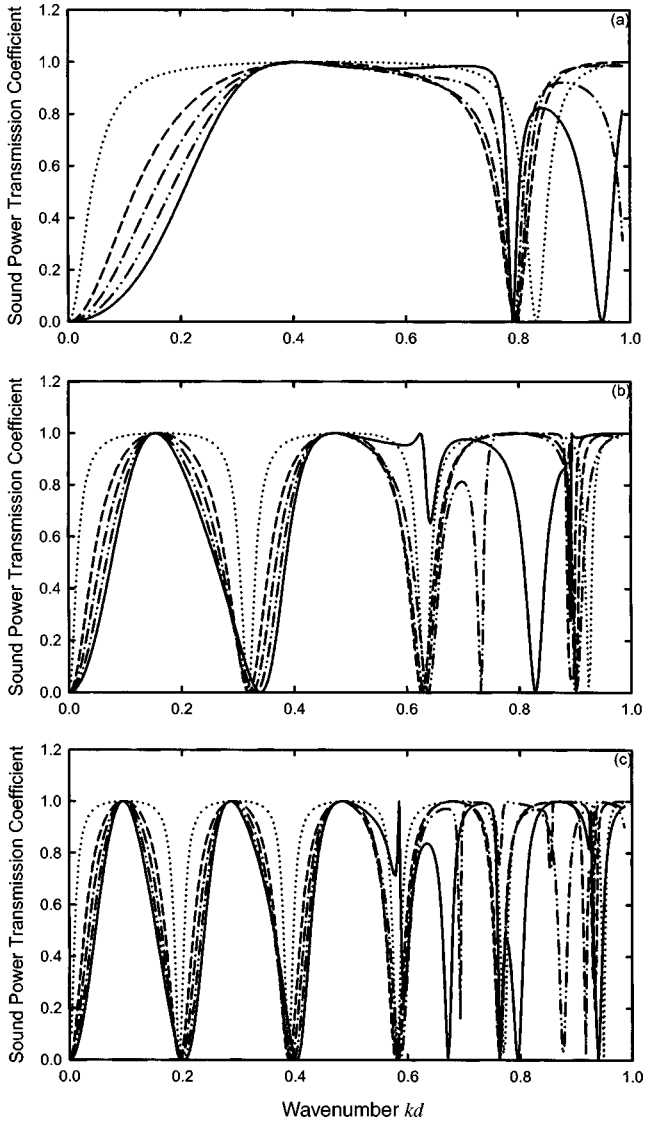


FIG. 7. Sound power transmission across open branches. $l/d=$ (a) 1; (b) 3; (c) 5. · · · ·: $w/d=0.3$; — — —: $w/d=0.9$; —·—·—·—: $w/d=1.2$; –··–: $w/d=1.5$; ——: $w/d=1.8$.

the ''both ends opened'' and ''one end closed one end opened,'' can be defined according to the wave patterns inside the sidebranches.[1] The former refers to the case where low-pressure regions are found at the two ends of the branch. The latter denotes the situation where the high-pressure regions are established at the duct wall opposite the branch.

The sound power transmission coefficients shown in Figs. 7(a) (b), and (c) suggest that these branches are high-pass filters at low frequency. However, one can see from the same figures that the branch length l this time has significant effects on the sound transmission. For short branch length, for instance at $l/d=1$, one can find more broadband high-pass action at reduced branch width [Fig. 7(a)]. Also, apart from the case of vanishing kd, there is only one frequency at which no sound can go across the branch for $w/d<1$. This vanishing sound transmission occurs twice for $w/d>1.2$ at $kd<\pi$. The frequency of the first vanishing τ decreases as w increases and the rate of decrease decreases with increasing w. This is mainly due to a ''both-end-opened'' type of longitudinal branch resonance as shown in Fig. 8(a) (low pressure at the two ends of the branch). This appears consistent with the experimental results of Tang and Li.[6] However, as w increases, high sound energy accumulation inside the Tee-junction especially at $y=0$ is found at $w/d=1.8$ under the ''both-end-opened'' branch resonance [Fig. 8(b)]. The second vanishing τ is due to the resonance of the first asymmetric opened branch mode [Fig. 8(c)].

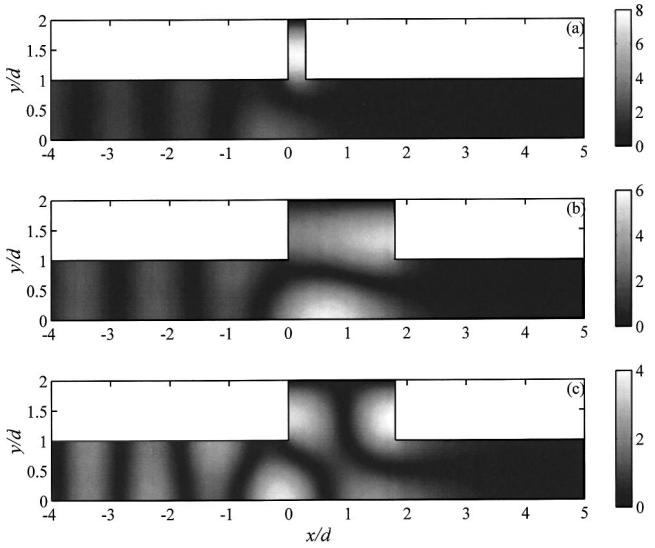


FIG. 8. Sound fields at frequencies of vanishing sound transmission for open branch cases. (a) $kd=0.831\pi$, $w/d=0.3$; (b) $kd=0.789\pi$, $w/d=1.8$; (c) $kd=0.949\pi$, $w/d=1.8$. $l/d=1$.

The increase in the branch length l results in more occurrence of branch mode resonances and thus more frequencies of vanishing sound transmission [Figs. 7(b) and (c)]. For $l/d=3$, the vanishing τ observed at $kd\sim0.32\pi$ and 0.64π are due to the type of planar longitudinal branch resonance illustrated in Figs. 8(a) and (b). However, the nodal and antinodal planes become more inclined at large w [Fig. 9(a)] and at $w/d=1.8$, and this resonance cannot take place because of the excitation of the first asymmetric standing branch mode [Fig. 9(b)]. Such coexcitation of the axial and asymmetric branch modes results in a vanishing pressure region at the entry of the branch. The effective branch length is greatly reduced, increasing the sound transmission inside the main duct and thus the rise in sound transmission at the corresponding frequency observed in Fig. 7(b). The vanishing τ for $w/d>1$ at $kd>0.7\pi$ is due to the excitation of higher-standing branch modes. This pattern repeats as kd approaches π. The situations for $l/d=5$ are the same except that the variation of τ with frequency is more rapid. Thus, they are not discussed further. The observed high sound transmission before the excitation of higher-branch mode resonance for all w and l is due to a ''one-end-closed-one-end-opened'' type branch resonance[1] with high pressure buildup around the entry of the branch. A typical example of such a sound field is given in Fig. 9(c). This is consistent with the experimental observation of Tang and Li.[6]

Two length corrections are involved in this opened-branch case. One is the acoustic length correction of the main duct section upstream of the sidebranch, and the other that of the sidebranch. Figure 10(a) shows three examples of

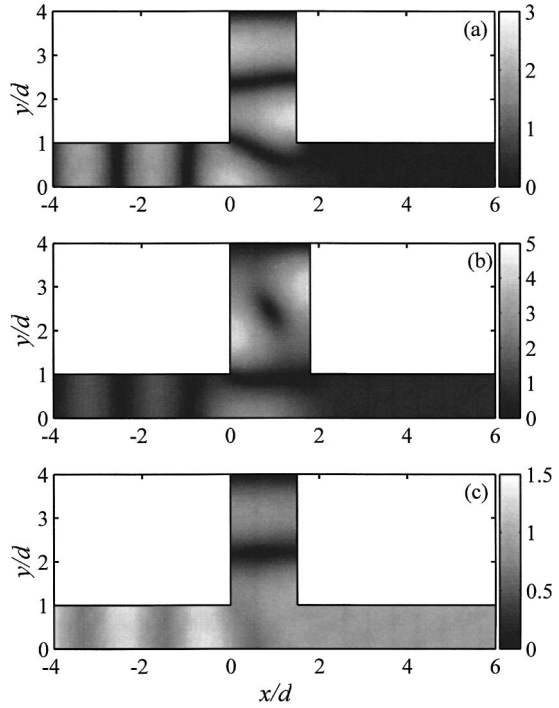


FIG. 9. Some examples of resonance inside opened sidebranch with $l/d=3$. (a) $kd=0.630\pi$, $w/d=1.5$; (b) $kd=0.643\pi$, $w/d=1.8$; (c) $kd=0.551\pi$, $w/d=1.5$.

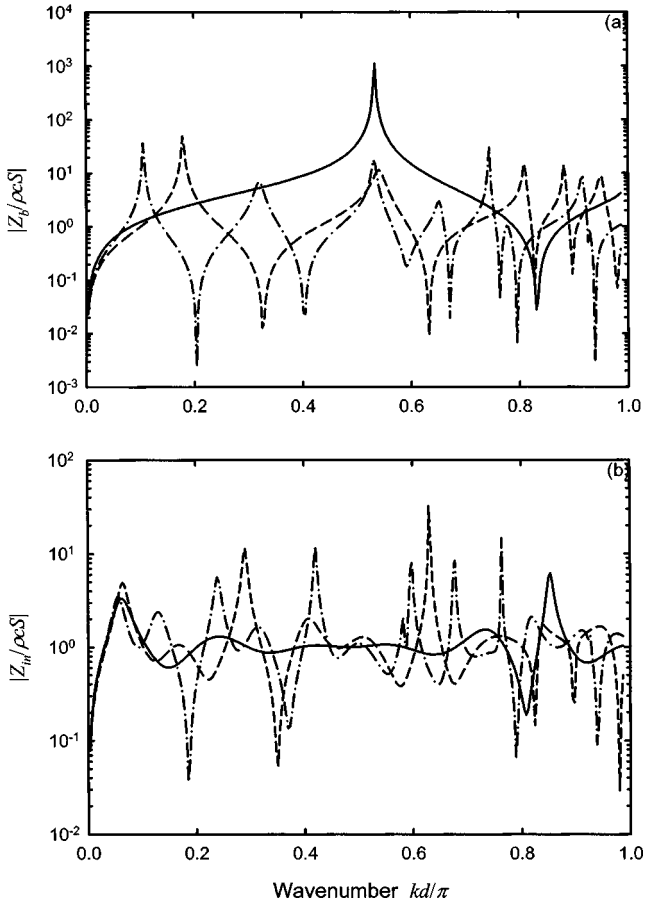


FIG. 10. Variation of mechanical impedance magnitudes with wave number for open-branch cases. (a) Sidebranch; (b) input impedance of main duct. ——: $l/d=1$, $w/d=0.3$; — —: $l/d=3$, $w/d=1.3$; —·—·—·—: $l/d=5$, $w/d=1.8$.

the sidebranch impedance Z_b calculated using the formula in Ref. 1. The pattern of impedance variation of the $l/d=1$, $w/d=0.3$ case appears to be very similar to that of an open-end tube.[1] The situations with $l/d=3$, $w/d=1.3$ and $l/d=5$, $w/d=1.8$ are similar, except that there are a number of higher-branch standing wave modes being excited as discussed before. The weak impedance is due to a ''both-end-opened'' type branch resonance discussed before. The cases of high resistance (branch impedance magnitude as well) are due to the ''one-end-closed-one-end-opened'' branch resonance [Fig. 9(c)]. The corresponding input impedance Z_{in} is shown in Fig. 10(b). One can observe similar frequency dependence as for Z_b at low frequencies only. Z_{in} depends very much on the impedance produced by the branch, such that the resonance does not occur so regularly as in the case of infinitely long branches (Fig. 5). There are lots of vanishing reactance cases (not shown here), but only those associated with a weak impedance magnitude and a relatively weak total acoustic pressure at the inlet $(x/d=-5)$ indicate the occurrence of resonance. One should also note that Z_{in} depends on L, but the corrections do not.

The branch length corrections of the initial few branch resonance are summarized in Fig. 11(a). Those related to frequencies where higher branch modes are excited are excluded as the formula of Ref. 1 may not apply directly to such situation. It is noticed that for the high-sound-transmission cases (large branch resistance), the corrections are negative and do not vary much with branch length and the order of the resonance. For the case of vanishing sound transmission, the corrections are positive. They decrease as branch length increases, but are reduced at increased resonance order.

The acoustic length correction in the main duct does not show a very well defined trend of variation with branch length, branch width, or the order of resonance [Fig. 11(b)], but it is certain that the branch resonance plays an important role in it. For $l/d=1$, ϵ_1 decreases with increasing w, while the other corrections show the opposite trend. The resonance in the main duct in this case, especially the first and second ones, does not depend very much on the branch resonance as the first important branch resonance which results in serious reflection is located at kd near to π. The weak reflection, especially for short w [Fig. 10(b)], allows easy passage of sound across the branch and thus the large length correction observed. As the order of resonance increases, the duct section upstream of the branch resonates with a shorter wavelength and the correction increases with w, but decreases as the order of resonance increases. Only four duct resonance are observed at $kd<\pi$ for $l/d=1$.

The situation with $l/d=3$ appears very different from that with $l/d=1$. The first branch resonance causing vanishing sound transmission takes place at $kd\sim0.33\pi$, the second one being at $kd\sim0.67\pi$. The strong reflection around these resonance frequencies gives rise to duct section resonance. The first two appear around $kd\sim0.3\pi$ and the third and fourth orders around $kd\sim0.6\pi$ [Fig. 10(b)]. The first and third duct resonances are due to the ''both-end-opened'' branch resonance, while the other two are produced by the

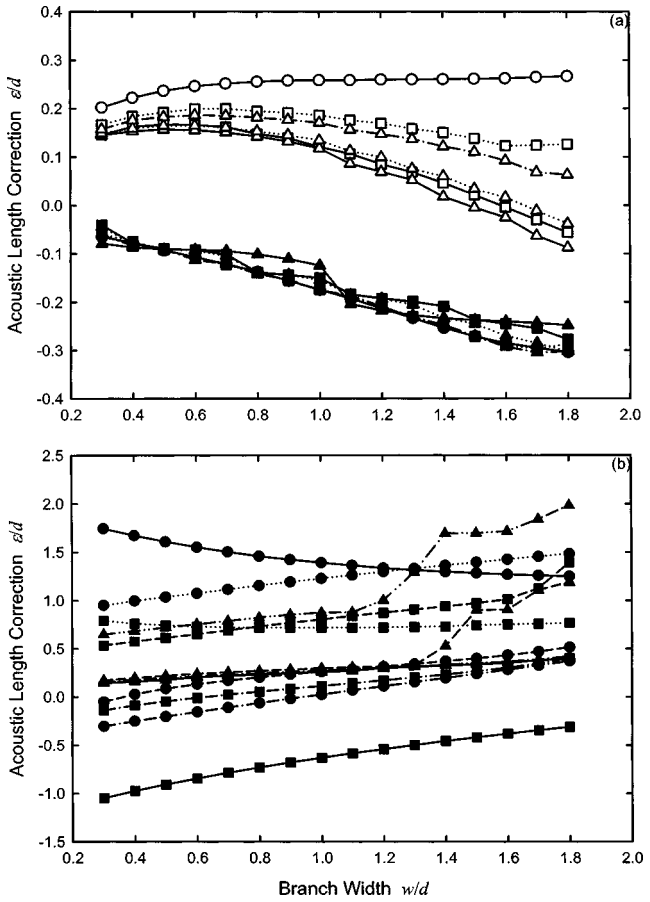


FIG. 11. Acoustic length corrections for opened sidebranches. (a) Branch length correction; (b) length correction of upstream duct section. ●: $l/d=1$; ■: $l/d=3$; ▲: $l/d=5$. ——: first order; · · · ·: second order; —·—·—·—: third order; — — —: fourth order; –··–: fifth order. Closed symbols: weak impedance; open symbols: strong impedance.

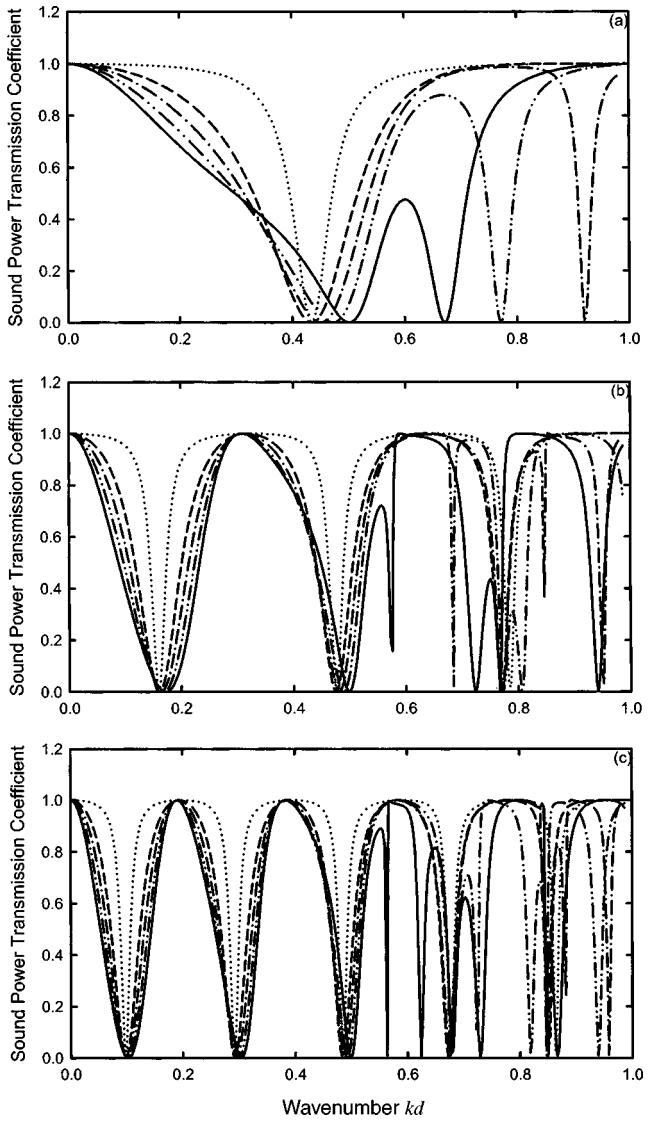


FIG. 12. Sound power transmission across closed sidebranches. $l/d=$ (a) 1; (b) 3; (c) 5. · · · ·: $w/d=0.3$; — — —: $w/d=0.9$; —·—·—·—: $w/d=1.2$; –··–: $w/d=1.5$; ——: $w/d=1.8$.

''one-end-opened-one-end-closed'' branch resonance. In the former case, the strong reflection takes place before the wave goes into the Tee-junction, resulting in a negative length correction. The opposite occurs in the latter cases. Examples of such resonance can be found in Figs. 9(a) and (c). For this branch length, the length corrections increase with w except for that of the second order resonance, which shows a fairly constant correction regardless of the value of w.

The very good correlation between the magnitudes of Z_b and Z_{in} for $l/d=5$ shown in Fig. 10 suggests that the corresponding duct resonance is mainly controlled by the ''both-end-opened'' branch resonance, the first one of which occurs at $kd\sim 0.2\pi$. Similar to the case of an infinitely long branch, the length corrections for the first three resonance are very close to each together regardless of the branch width, and all corrections are positive. The magnitudes of the corrections of the first three resonance are also comparable to those of the infinitely long branches (Fig. 6). Those of the fourth resonance appear close to those of the first three resonance until the higher branch mode is excited. As discussed before, the higher asymmetric branch mode helps the spreading of acoustical energy across the junction, resulting in a more rapidly increase in the length correction with branch width. Fifth order resonance is observed in this case. Early rapid rise in the length correction is observed. This is consistent with the results for the infinitely long branch. In fact, a similar rapid rise in length correction can also be found in the fourth order resonance of the $l/d=3$ case. Results shown in Fig. 11(b) suggest that the duct length corrections can be comparable to d, implying that the conjecture of Tang and Li[6] on large branch separation correction is possible.

C. Closed branches

The closed branches produce a low-pass filtering effect in sound propagation such that $\tau\rightarrow 1$ as $kd\rightarrow 0$. Two types of branch resonance will occur. The first one is the ''both-end-closed'' one where high pressure is created in the junction. The second one is again of the ''one-end-closed-one-end-opened'' type but this time the low-pressure region is found inside the junction.

Figures 12(a), (b), and (c) illustrate the frequency variations of the sound power transmission coefficients for $l/d=1$, 3, and 5, respectively, at various w. While the wave number at the second vanishing sound transmission de-

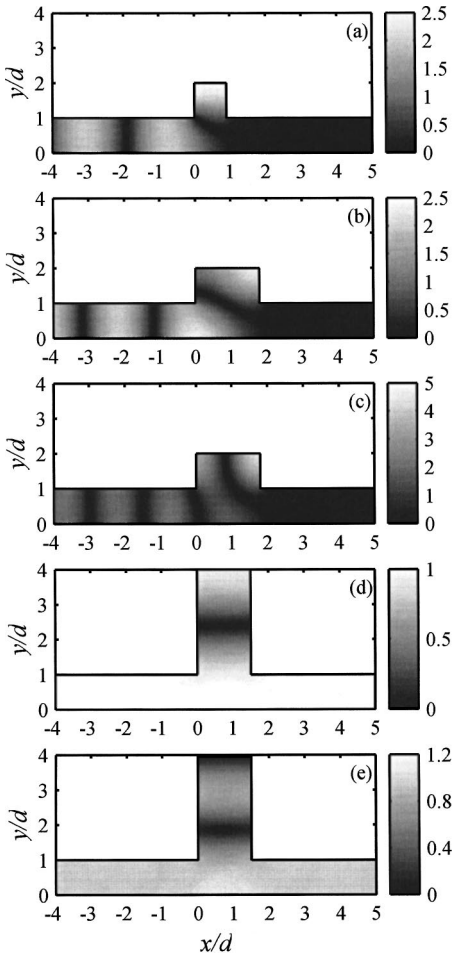


FIG. 13. Some typical examples of resonance inside the closed sidebranches. (a) $kd=0.433\pi$, $w/d=0.9$, $l/d=1$; (b) $kd=0.500\pi$, $w/d=1.8$; $l/d=1$; (c) $kd=0.672\pi$, $w/d=1.8$; $l/d=1$; (d) $kd=0.312\pi$, $w/d=1.5$; $l/d=3$; (e) $kd=0.649\pi$, $w/d=1.5$; $l/d=3$.

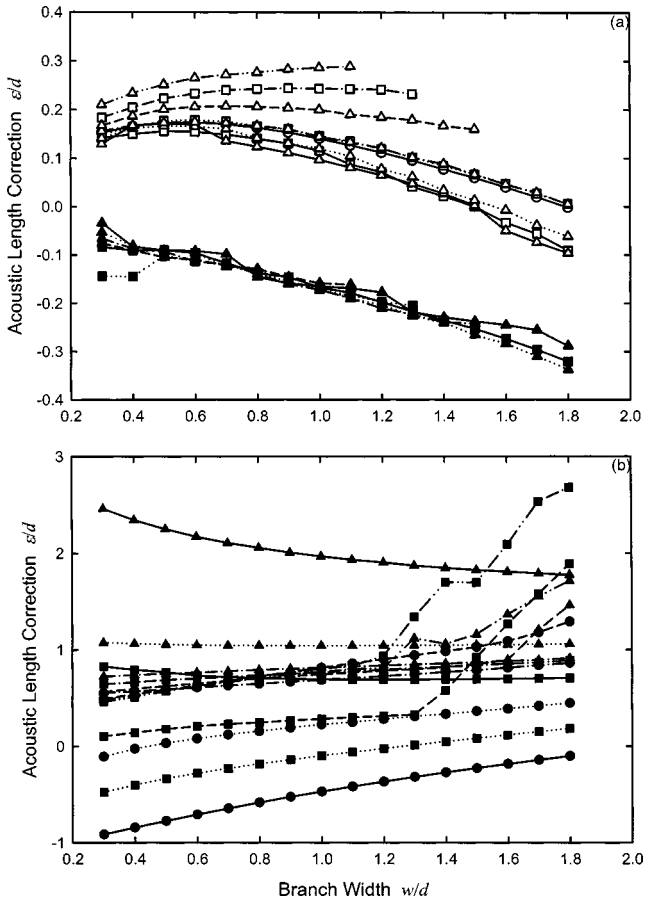


FIG. 14. Acoustic length corrections for closed sidebranches. (a) Branch length correction; (b) length correction of upstream duct section. Legends: same as those of Fig. 11.

creases with increasing w as in the opened-branch cases, that of the first one increases as w increases as shown in Fig. 12. Again the first vanishing sound transmission is due to a planar longitudinal branch resonance, but this time this resonance results in high acoustic pressure magnitude at the closed-branch end and a low acoustic pressure region inside the Tee-junction [Fig. 13(a)]. This resonance resembles that of the ''one-end-closed-one-end-opened'' tube. As w increases, it is more difficult to sustain a planar standing wave form inside the branch, as illustrated in Fig. 13(b), as the boundary condition at branch end in this case does not enforce uniform pressure there. The branch length for the resonance decreases as w increases, resulting in a shift of the first vanishing sound transmission kd toward the higher side. The nodal plane eventually intersects with the branch end and it is then obvious that the excitation of the first asymmetric standing branch mode soon follows, resulting in an earlier occurrence of the second vanishing sound transmission at lower kd [Fig. 13(c)].

As the branch length increases, more branch resonance occurs at lower kd as in the opened branch cases [Figs. 12(b) and (c)]. The longer the branch, the larger the number of plane wave resonances inside the branch before the excitation of the higher-standing branch modes. The patterns of τ variations in the range $0.5\pi<kd<0.6\pi$ for $w/d=1.8$, $l/d=3$, in the range $0.8\pi<kd<0.9\pi$ for $w/d=1.5$, $l/d=3$ [Fig. 12(b)] and in similar ranges for the $l/d=5$ cases [Fig. 12(c)] and some others are due to the same mechanism as described for the $l/d=1$ cases [Figs. 13(a)–(c)]. Thus, the corresponding results are not discussed further.

Comparing the results shown in Figs. 7 and 12, one can find that the frequencies for the vanishing sound transmission due to the longitudinal plane wave branch resonance for the opened-branch cases correspond approximately to those of ideal sound transmission for the closed-branch cases, and vice versa. However, in the closed-branch cases, the strong and vanishing transmissions are due to a ''both-ends-closed'' type and a ''one-end-closed-one-end-opened'' type branch resonance, respectively. Some examples of the corresponding sound fields are given in Figs. 13(a), (d), and (e). The large accumulation of acoustical energy on the main duct wall immediately opposite the branch entry illustrated in Figs. 13(d) and (e) tends to suggest a relatively large branch length correction.

Figure 14(a) summarizes the branch length corrections for the first few closed-branch resonances. The observations are similar to those for the opened sidebranches [Fig. 11(a)]. The corrections correspond to the strong sound transmission cases, this time due to the ''both-end-closed'' type branch resonance, are negative and are not sensitive to the order of resonance and the branch length before the higher branch

modes are excited. Their values are also comparable to those of the opened-branch cases. The corrections associated with the vanishing sound transmission, resulted from the ''one-end-closed-one-end-opened'' type branch resonance, increase with the order of resonance, but unlike the case of the opened sidebranch, they increase with branch length. The data obtained after the excitation of higher branch modes are again excluded.

Figure 14(b) illustrates the length corrections in the main duct upstream section. Again, these resonance correspond to weak input impedance and are very often associated with the ''one-end-closed-one-end-opened'' branch resonance. The branch resonance is even more important when the first asymmetric branch mode is excited as it forces a weak pressure in the middle of the junction, causing strong upstream reflection [for instance, see Fig. 13(b)]. One can expect that the branch resonance takes place at frequencies earlier than those in the opened branch cases. For $l/d=1$, negative corrections are observed for the first order resonance over the range of w investigated. Then the corrections increase monotonically with the order of resonance. Excitation of higher branch modes is illustrated by the results of the fourth order resonance. The corrections increase with w, but show a tendency to converge before they are affected by the higher branch modes.

The situations at $l/d=3$ resemble very much those in the opened-branch case, although the variations of the corrections with the order of resonance are opposite. In this case, the corrections due to the odd order resonance are positive and the others negative. Together with the results for the opened branches, one can conclude that the correction is negative and positive whenever a high- and low-pressure region is enforced by the branch resonance respectively. Rapid increase in the corrections can be observed again after the excitation of the higher branch modes. The duct resonance is again controlled by the vanishing sound transmission branch resonance when $l/d=5$ (not shown here). However, before the occurrence of the higher branch mode resonance, the associated corrections decrease with some degrees of convergence as the order of resonance increases. The positive corrections are expected as the branch resonance this time tends to force a low-pressure region within the junction. One can again observe the rapid rise in the length corrections after the higher branch modes are excited.

IV. CONCLUSIONS

In the present study, the sound transmission characteristics of and the acoustic length corrections due to a Tee-junction along an infinitely long duct are investigated numerically using the method of finite elements. The sidebranches considered here are the infinitely long branches, and opened and closed branches. The frequency range of the present study is up to the first eigenfrequency of the main duct.

For infinitely long sidebranches, a sharp drop in the sound transmission coefficient is found at a frequency slightly higher that the first asymmetric branch eigenfrequency. The increase of the sound transmission effectiveness resumes afterward. The relatively uniform particle velocity fields at the entry of the sidebranches at the higher branch mode eigenfrequencies delay such higher-mode excitations. The corresponding acoustic length corrections of the upstream duct section increase with branch width and the order of the resonance.

Opened sidebranches produce alternate strong and weak sound transmissions due to the occurrence of the planar branch mode resonance. At strong sound transmission, the branch length corrections are negative and their magnitudes increase as the branch width increases. They do not really depend on the order of the branch resonance. For the other planar branch resonance cases, the corrections show a slight increase with short branch width and then decrease as the branch width increases beyond that of the main duct, except for the first order resonance, which appears to stay at a particular value as the branch width increases. They also decrease with branch length and the order of the resonance. The length corrections related to the main duct do not show a very general trend, but the magnitudes of the corrections show a tendency to converge at increased branch width before the nonplanar branch modes are excited. After the excitation of such modes, the corrections increase rapidly and keep on increasing as the branch width increases.

The closed sidebranches produce actions similar to that of an expansion chamber before the nonplanar higher branch modes are excited. The excitation of the nonplanar higher branch modes reduces the sound transmission coefficients. The branch length corrections due to the strong sound transmission branch resonance are negative and the values are comparable to those of the opened-sidebranch cases. The corresponding length corrections at vanishing sound transmission increase with branch length and the order of resonance. The duct length corrections show features that are in general opposite to those observed in the opened-sidebranch cases. However, their magnitudes still show some degrees of convergence at increased branch width unless a nonplanar higher branch mode is excited. These excitations result in rapid increases in the correction magnitudes.

ACKNOWLEDGMENT

This work is supported by a grant from the Research Committee, The Hong Kong Polytechnic University (Project No. G-YD59).

[1] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 4th ed. (Wiley, New York, 2000).
[2] D. D. Reynolds and J. M. Bledsoe, *Algorithms for HVAC Acoustics* (American Society of Heating, Refrigeration and Air-Conditioning Engineers, Atlanta, Georgia, 1991).
[3] J. W. Miles, ''Diffraction of sound due to right angled joints in rectangular ducts,'' J. Acoust. Soc. Am. **19**, 572–579 (1947).
[4] J. C. Bruggeman, ''The propagation of low-frequency sound in a two-dimensional duct system with T joints and right angle bends: Theory and experiment,'' J. Acoust. Soc. Am. **82**, 1045–1051 (1987).
[5] T. C. Redmore and K. A. Mulholland, ''The application of mode coupling theory to the transmission of sound in the sidebranch of a rectangular duct system,'' J. Sound Vib. **85**, 323–331 (1982).
[6] S. K. Tang and F. Y. C. Li, ''On low frequency sound transmission loss of

double side-branches: a comparison between theory and experiment,'' J. Acoust. Soc. Am. **113**, 3215–3225 (2003).
[7] S. K. Tang and C. K. Lau, ''Sound transmission across a smooth nonuniform section in an infinitely long duct,'' J. Acoust. Soc. Am. **112**, 2602–2611 (2002).
[8] P. M. Radavich and A. Selamet, ''A computational approach for flow-acoustics coupling in closed side branches,'' J. Acoust. Soc. Am. **109**, 1343–1353 (2001).
[9] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1962).

Application of theoretical modeling to multichannel active control of cooling fan noise[a)]

Kent L. Gee[b)] and Scott D. Sommerfeldt
Department of Physics and Astronomy, N-283 ESC Brigham Young University, Provo, Utah 84602

(Received 3 April 2003; revised 11 October 2003; accepted 20 October 2003)

Multichannel active control has been applied to the global reduction of tonal noise from a cooling fan. In order to achieve consistent far-field attenuation of multiple harmonics of the blade passage frequency (BPF) of the fan, an analytical model has been applied to the control system in order to determine appropriate transducer configurations. The results of the modeling show that the additional global reduction possible by locating acoustically compact secondary sources coplanar with a compact primary source rapidly lessens as the number of symmetrically placed sources is increased beyond three. Furthermore, the model suggests that there are locations in the extreme near field of the sources that can be considered ideal for the minimization of far-field radiated power. Experiments carried out show that a four-channel control system is more effective than a two-channel system at achieving far-field attenuations, especially at the higher harmonics of the BPF for the fan tested. In addition, greater far-field mean-square pressure attenuations are achieved with the error microphones located along the calculated ideal regions than for nonideal placement.
 [DOI: 10.1121/1.1631940]

PACS numbers: 43.50.Ki [KAC] Pages: 228–236

I. INTRODUCTION

Fan noise has been a widely researched application of active noise control (ANC) since the development of adequate digital signal processing technology. Although the vast majority of investigations has been directed toward ducted fans,[1–4] such as turbofans,[2–4] a small number of studies has been carried out on the reduction of radiated tonal noise from the axial flow fans used to cool electronic devices.[5–7] Single-channel efforts, by Quinlan,[5] Lauchle *et al.*,[6] and Wu,[7] have shown that active control is indeed a viable method of reducing harmonics of the blade passage frequency (BPF), which typically dominate the overall spectrum. However, at the same time, the research carried out previously has suggested the need for further study. Two specific improvements that are examined in this paper are to (1) increase global reductions of the first four harmonics of the BPF by using a multichannel control system, and (2) increase practicality of the ANC system by locating error sensors in the near field of the fan. Because near-field placement of error sensors while seeking far-field attenuation has been deemed ''risky'' by Hansen and colleagues,[8,9] an optimization technique is needed to select an appropriate control actuator arrangement and then to ensure the sensors' locations result in consistent far-field reductions for that actuator configuration.

Performance optimization of an ANC system depends on a number of factors, including the characteristics of the primary noise source, noise environment, and performance criteria. A number of studies have been performed with regards to the maximization of global ANC in a variety of contexts. While the majority of adaptive active control schemes have sought to minimize squared pressure using a form of the filtered-*x* algorithm, other methods have been studied. In some situations, a cost function based on energy density[10] or mean active intensity[8] has been found to be superior to the squared pressure cost function. Much of the research aimed at optimizing transducer configurations has primarily dealt with the reduction of structural acoustic radiation,[11–13] and, more specifically, the noise generated by extended radiators such as electrical transformers.[14,15] Because a small cooling fan represents a significantly different noise source, much of the research has not been found to be extremely applicable and therefore is only briefly summarized where appropriate.

The research regarding the role of transducers in ANC can be largely grouped in two categories, namely investigations involving control actuators and error sensors. Nelson *et al.*[16] described the minimum power output for various configurations of free-field point sources, thereby showing the maximum sound power attenuation possible for similar configurations consisting of real monopole-type primary and secondary sources. Martin and Roure modeled the primary source field using a spherical harmonic multipole expansion in order to determine the optimal location for control sources.[14,17] Others have also examined the use of multipoles as control sources because significant global far-field attenuation may still be achieved with a primary source and secondary multipole located at a relatively large fraction of a wavelength from each other.[18–20]

The optimization of error sensing techniques has also received considerable attention. A number of near-field error sensing schemes has been theoretically examined by Qiu *et al.*[8] Despite their finding that minimization of the mean active intensity normal to a surface surrounding all sources is

[a)]Portions of this work were presented in ''Multi-channel active control of cooling fan noise,'' *143rd Meeting: Acoustical Society of America*, Pittsburgh, PA, June 2002, and ''Multichannel active control of cooling fan noise,'' *Proceedings of Inter-Noise 2002*, Dearborn, MI, August 2002.

[b)]Graduate Program in Acoustics, Pennsylvania State University, P.O. Box 30, State College, PA 16804. Electronic mail: kentgee@psu.edu

the ideal near-field sensing strategy, Berry, *et al.*[12] compared the minimization of intensity versus squared pressure for a vibrating plate and found that intensity minimization provided negligible improvement over near-field squared pressure minimization. Also, Wang[13] and Martin and Roure[17] made use of genetic search algorithms in order to determine the optimal locations for farfield error microphone locations.

In this paper, we discuss a modeling technique based on the work of Nelson *et al.*[16] is discussed which is used to both determine an appropriate configuration of control sources and to suggest near-field error microphone locations that result in consistent far-field reductions for multiple harmonics of the BPF radiated from a small cooling fan. Following a description of the modeling performed, experimental results are presented and analyzed.

II. THEORETICAL MODELING

A. Source modeling

1. Primary source

Before the development of an appropriate method to determine control system transducer locations, it is necessary to first discuss the disturbance noise source itself. The spectrum of a cooling fan consists of a broadband component upon which is superimposed one or more discrete-frequency peaks, which are harmonically related to the BPF. The dominant noise generation mechanism for the tonal noise with subsonic blade tip speeds is spatially unsteady loading on the fan blades.[21] Unsteady loading may be attributed to flow obstructions near the inlet or exhaust of the fan, such as fan support struts or finger guards.[22] These fluctuating lift and drag forces on the blades generate dipole-like noise. If the tonal frequency and fan diameter are such that the fan may be considered aeroacoustically compact, then the far-field radiation will be that of a dipole, often skewed, as measurements on unbaffled cooling fans have shown.[5,6] However, with the fan baffled, the radiation is closer to that of a monopole. At higher harmonics of the BPF, where the condition of compactness no longer holds, the far-field radiation pattern becomes increasingly complex and directional. The number of harmonics present, their amplitudes, and directional characteristics will largely depend on the nature of the unsteady loading.

Because the radiation at the BPF is fairly smooth and omnidirectional for the fan and experimental configuration used, the fan is modeled as a point monopole source, mounted in an infinite baffle. Although the experimental fan may not strictly be considered baffled, since it is mounted in an aluminum enclosure roughly the size of a personal computer chassis, the infinite baffle assumption greatly simplifies the calculations. Furthermore, even though the fan's radiation becomes more complex as a function of frequency, the decreasing wavelength causes the fan to appear more baffled at higher harmonics of the BPF. Because of this, the radiation of the fan is modeled as a monopole source up to the fourth harmonic of the BPF, which is the highest tone targeted by the ANC system.

2. Control sources

A practical ANC system that could be applied to typical commercial cooling fan applications (e.g., a personal computer) would have to be compact and self-contained, that is to say, the actuators and sensors would need to be mounted in or on the device. It is necessary, therefore, to determine actuator and sensor locations in the near field of the fan that lead to global far-field reductions. Because the fan is assumed to be baffled, the secondary sources will be located on the baffle. Furthermore, although the use of multipoles as control sources has been found to be advantageous in some instances,[14,17–20] the requirement of system compactness precludes the use of these types of actuators. Finally, because the loudspeakers used in this investigation as control actuators have a smaller diameter than the fan and can be considered more acoustically compact over the frequency range of interest, they are also modeled as point monopoles.

B. Determination of control source configurations

With the fan and ANC system actuators modeled as point monopole sources, an analysis similar to that of Nelson *et al.* can be performed in order to discover an appropriate configuration for control sources. The purpose of the approach is to determine the minimum power radiated by a given control source configuration. By examining various control source arrangements, the optimal configuration can then be determined. The procedure for finding the minimum radiated power for a system of point sources is well documented in Refs. 16 and 23, and therefore is briefly summarized as follows. First, control source locations are determined and the sources given arbitrary complex strengths relative to that of the primary source. Second, because the mutual acoustic coupling of multiple point sources is well understood analytically, a mathematical expression for the radiated power from all sources may then be obtained. Third, the source strengths that minimize the total radiated power may be obtained by differentiating the total radiated power expression with respect to the real and imaginary parts of each of the control source strengths, setting each resultant equation equal to zero, and solving for the optimal strengths. Other than the baffled source assumption, the analysis is only different from that of Nelson *et al.*, in the sense that all sources are required to be located in the same plane, a restriction not imposed by the previous research. The fact that the sources are baffled only scales all source strengths by a factor of 2; the relative strengths of primary and secondary sources and radiated power are identical to the unbaffled case.

The source strength optimization technique has been carried out for a number of configurations not included in the Nelson *et al.* analysis.[24] Two such configurations, that of four and eight secondary sources arranged symmetrically around and in the same plane as the primary source, are displayed schematically after the manner of Nelson and colleagues in Fig. 1. The minimum radiated power, relative to the primary source alone, is shown in Fig. 2 as a function of *kd*, which is the nondimensionalized separation distance between the primary and secondary sources. Displayed is the radiated power

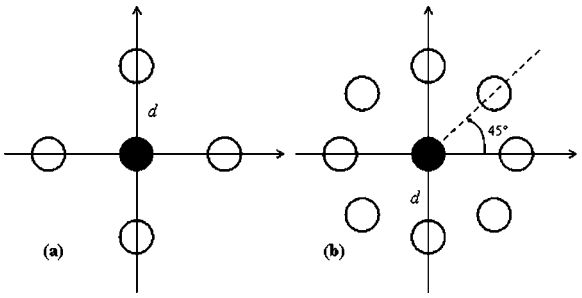


FIG. 1. Point source monopole configurations, after the manner of those shown by Nelson *et al.* in Fig. 5 of Ref. 16. The primary source (shaded) is located at the origin. Configuration (a) consists of four secondary sources, each at a distance, *d*, from the primary source and placed at 90° increments. Arrangement (b) contains eight secondary sources, each at a distance, *d*, from the primary source and are each separated by 45° angles.

for the configurations analyzed previously by Nelson *et al.*,[16] denoted by (N) in the figure legend, as well as for the source arrangements in Fig. 1. The source configuration that yields the most attenuation for a separation distance of less than a half-wavelength ($kd<\pi$) is the tetrahedral arrangement shown in Fig. 5(c) of Ref. 16, in which the primary source is at the center of the tetrahedron and equidistant from all four secondary sources. While this is not a valid arrangement under the imposed limitation that all sources be located in the same plane, it does demonstrate that greater suppression of power radiated from the primary source can be achieved with a three-dimensional source arrangement.

An examination of the five source configurations in which the limitation is met yields another important result: the additional reduction that can be achieved as more sources are added and symmetry is maintained rapidly lessens. As *kd* becomes small, three, four, and eight symmetrically arranged and optimized secondary sources provide essentially the same power reduction. While the greatest difference between the three and four secondary source arrangements for $kd<\pi$ is on the order of 1 dB, the largest difference between four and eight sources over the same range is less than 0.2 dB. For $kd>\pi$, even though the overall reduction possible for any configuration is small, it is interesting to note that for the range shown, the reduction in radiated power for the tetrahedral distribution is actually less than that of the four and eight coplanar source arrangements. However, because of the desire to locate the primary and control sources as close as possible, this latter result is considered secondary. The power-minimization results for the planar configurations described suggest that either three or four monopole-like actuators placed symmetrically around the fan is a practical limit to optimize global control of a baffled compact source.

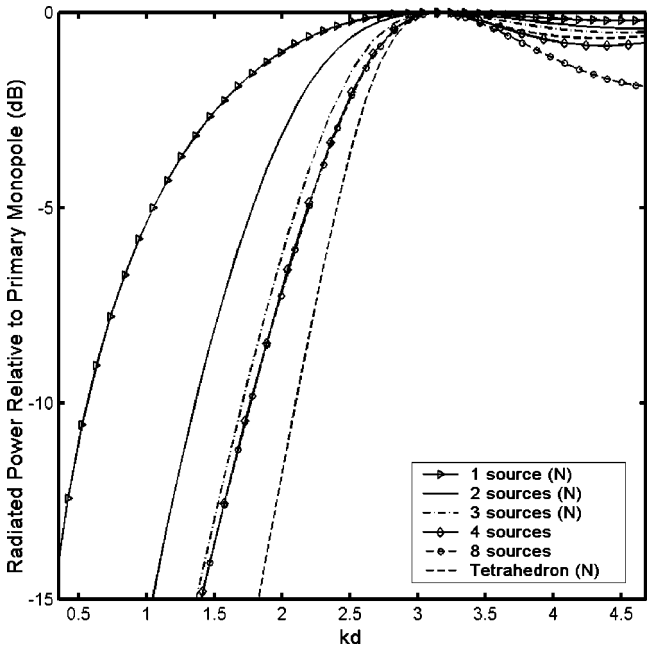


FIG. 2. Radiated power, relative to that radiated by the primary monopole itself, for secondary source arrangements. The secondary source strengths have been chosen so as to minimize the radiated power of the system. Those configurations marked by an (N) in the figure legend have been analyzed previously by Nelson *et al.* in Ref. 16. The other two arrangements are those depicted in Fig. 1.

C. Determination of near-field error sensor locations

The preceding analysis does not include the determination of error sensor locations that result in global control, which is a necessary step in developing a viable, robust ANC system. For this particular application, the issue is complicated by the requirement that the error sensors be located in the near field, yet yield far-field reductions. Near-field error sensor placement has been traditionally warned against, in that sound power attenuation can vary rapidly as a function of location[9] and far-field pressure reductions may be accompanied by pressure increases in the near field.[8] The subsequent discussion describes the determination of appropriate near-field error sensor locations based on a continuation of the analysis performed in the previous section.

For the active control of free-field radiation, the optimal error sensor locations are those where the acoustic pressure attenuation is greatest when the control sources' strengths are such that the overall power radiated is minimized.[9] Error sensors may then be suitably located in the near field by discovering the point or regions of greatest pressure attenuation. The method of locating these regions used in this investigation is that of plotting the two-dimensional magnitude of the near-field pressure attenuation in the plane of the sources for the condition of minimized power radiation, and then determining an appropriate location graphically. An alternative to the graphical method would be to develop an analytical expression for optimal placement as a function of frequency and position relative to the primary source, however, this approach was not utilized for two reasons. First, just as the fan and control loudspeakers are not point sources, the error sensors used in practice are not point sensors, but microphones with diaphragms of finite width, and therefore it would likely be pointless to attempt to ascertain with extreme precision the optimal locations. Second, it may be shown that some of the points considered ideal are so close to the primary source that, in practice, noise induced by the airflow across the microphones due to the fan results in a poor signal-to-noise ratio. For these reasons, the graphical method has been chosen as a guide for approximate near-field error microphone placement.

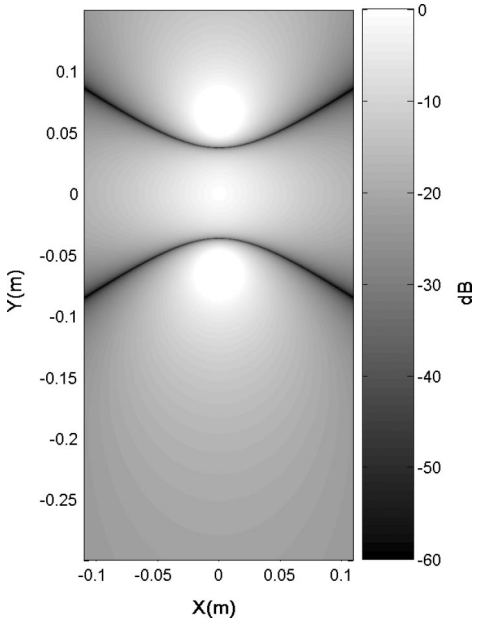


FIG. 3. The magnitude of the radiated pressure, relative to that of the primary source by itself, at 370 Hz (1×BPF) along the dimensions of the experimental enclosure for two optimized secondary monopoles located at (0,0.06) and (0,−0.06) m and the primary source located at the origin. The regions for optimal error sensor placement are demarcated by the pressure nulls.

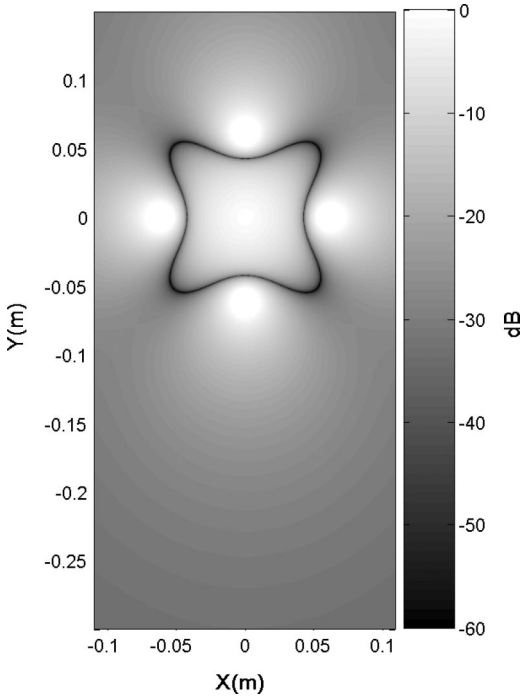


FIG. 4. The magnitude of the radiated pressure, relative to that of the primary source by itself, at 370 Hz (1×BPF) along the dimensions of the experimental enclosure for four optimized secondary monopoles located at (0.06,0), (0,0.06), (−0.06,0), and (0,−0.06) m. The primary source is located at the origin.

Although other control configurations have been studied both theoretically and experimentally,[24,25] the effect of error sensor location on global reductions for two source arrangements has received particular experimental focus. These configurations are that of the two sources, placed on opposite sides of the primary source, as discussed by Nelson *et al.*, and the four sources located symmetrically around the primary source. The magnitude of the near-field pressure, normalized by the pressure of the primary source by itself, for these two configurations is plotted in Figs. 3–5, for the dimensions of the enclosure face on which the control sources are located. For these figures, the control source strengths have been set so as to minimize the overall radiated power. The distance between the primary source located at the origin and the secondary sources is 0.06 m. Figures 3 and 4 display the pressure magnitude for 370 Hz, the BPF of the fan used in the experiments, for the two and four control source arrangements, respectively. It is clearly seen from these two plots of the near-field pressure that nulls exist in the radiation pattern when the control sources radiate in the minimized-power state. Because the pressure attenuation is the greatest in these regions, the ideal error sensor locations are delimited by the nulls. In the case of the two control source configuration, the two nulls that curve between the primary and secondary sources continue out to the far-field, approaching asymptotes that pass through the origin. The slope of these asymptotes gradually lessens as frequency increases; however, over the range plotted the arclength is small enough that the locations of the nulls remain fairly constant for the first four harmonics of the BPF.

The near-field pressure null for the four source configuration demonstrates markedly different behavior than those of the previous analysis. Instead of continuing out to the far-field, the null is of closed form and only exists in the near field. The frequency dependence of the null location is shown by comparing Figs. 4 and 5, the latter of which shows the near-field pressure attenuation at the fourth harmonic of

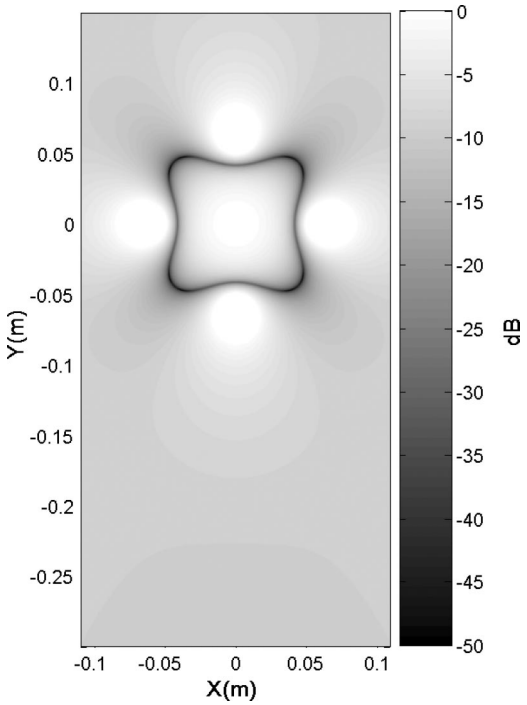


FIG. 5. The magnitude of the radiated pressure, relative to that of the primary source by itself, at 1480 Hz (4×BPF) for the same four optimized sources as in Fig. 4, showing the changed shape in the near-field null at 45° angles relative to the location of the secondary sources.

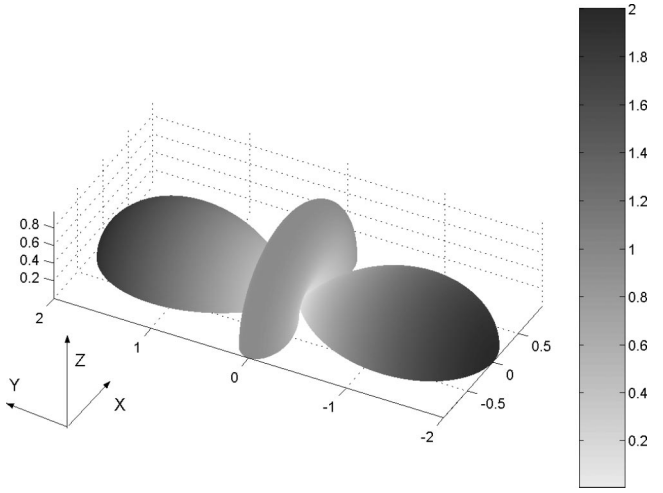


FIG. 6. Linear three-dimensional directivity at 370 Hz for the system with two optimized secondary sources described in Fig. 3. Symmetry of the radiation about the *y* axis signifies that the plane of the sources may be considered an ideal location for error sensors, because there is no dependence on the polar angle, θ.

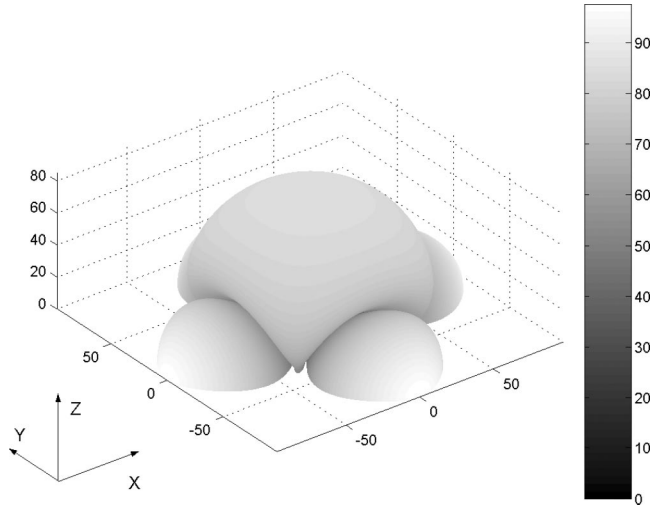


FIG. 7. Three-dimensional directivity, represented logarithmically in dB, at 370 Hz and $r=0.072$ m for the control configuration with four optimized secondary sources described in Fig. 4. The radiation is not symmetric about any of the three Cartesian axes in this case, demonstrating that the near-field null location is a function of both the azimuthal angle, ϕ, and the polar angle, θ.

the BPF, which is 1480 Hz. Like the two control source case, the dependence of the null location on frequency is not extreme, but it does suggest that an appropriate compromise would be to locate an error sensor closer to a control source rather than equidistant between two control sources, where the null location change is the greatest. This technique clearly demarcates appropriate near-field error sensor locations that should allow the control sources to operate under optimal source strength conditions, and therefore lead to maximum far-field attenuations.

D. Justification of error microphone placement in source plane

Before proceeding to an experimental verification of this technique for the two configurations analyzed, it is important to justify placement of the error sensors in the source plane, rather than off the plane. After all, it was shown in the previous section that to begin to truly maximize the reduction of far-field power for $kd<\pi$, a three-dimensional distribution of secondary sources is needed (e.g., the tetrahedral configuration of Nelson *et al.*). This justification of error sensor placement may be expressed as a need to determine the behavior of the pressure attenuation off the source plane to see if the nulls are perhaps deeper than in the plane of the sources. For the two control source case, it can be seen from the directivity plot in Fig. 6 that the pressure radiated from the system of sources is symmetric about the *y* axis. It is then clear that placement of the error sensors in the plane of the sources (the *x*-*y* plane in Fig. 6) may be considered ideal for this configuration. However, for the four control source case, the radiation is not axisymmetric, as can be easily seen in Fig. 7, which shows the directivity of the radiation for 370 Hz at a distance $r=0.072$ m from the primary source. The distance has been chosen on the basis that the three-dimensional directivity will include the near-field null at a number of locations in the plane of the sources, providing the opportunity to observe the behavior of the nulls off the plane.

The pressure directivity for the optimized four control source configuration as a function of azimuthal angle, ϕ, and polar angle, θ, is displayed in Fig. 8. The polar angle, θ, is measured from the axis perpendicular to the source plane, and the azimuthal angle, ϕ, is measured from the *x* axis, shown in Figs. 3–7. The directivity shows that the depth of the null is constant as θ, which is equal to $\pi/2$ in the plane of the sources (i.e., $z=0$), is lessened. This behavior of the null demonstrates that the plane of the sources may also be considered ideal for the placement of error sensors. This same behavior has been verified for various near-field radii (r) and frequencies and is found to be consistent as long as r is such that it includes the near-field null. If error sensors are to be placed anywhere beyond the null, it may also be shown that

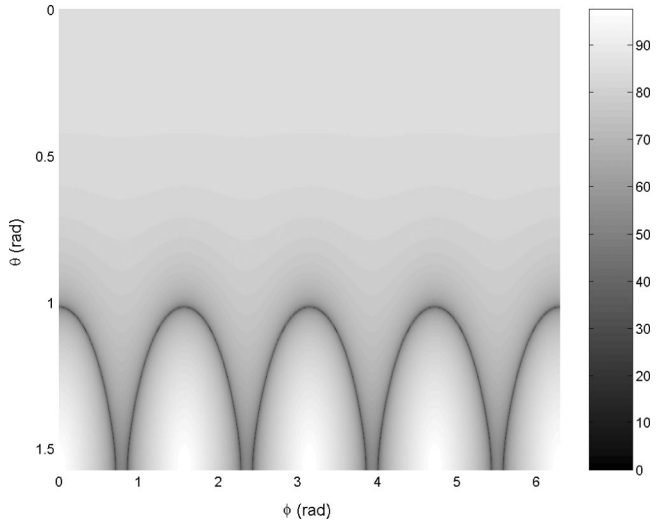


FIG. 8. A two-dimensional representation of Fig. 7, with the polar angle, θ, on the vertical axis and the azimuthal angle, ϕ, on the horizontal axis, showing the behavior of the pressure nulls off the source plane, where $\theta=\pi/2$.

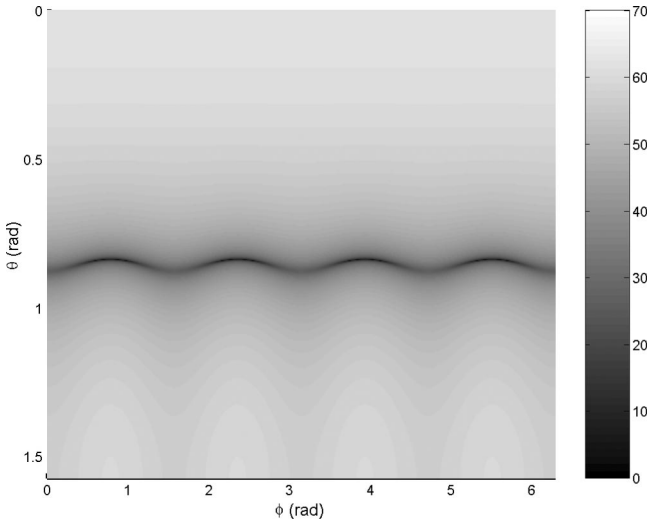


FIG. 9. Two-dimensional representation of the directivity at $r=1.0$ m and a frequency of 1480 Hz for the four secondary source arrangement, demonstrating that the source plane may no longer be considered an ideal location for error sensors.

the greatest areas of pressure attenuation are not found in the source plane, but above it (for example, see Fig. 9).

III. EXPERIMENTAL SETUP

The results of the theoretical modeling carried out indicate that there are appropriate near-field error sensor locations for the two and four control actuator configurations. These configurations are now examined experimentally. Although Ref. 25, which comprises a parallel investigation on the practical aspects of multichannel ANC of cooling fan noise, contains a more complete description of the experimental apparatus used, the major components are summarized here.

The cooling fan used is a seven-bladed, 80 mm DC Mechatronics fan, which has been housed in a $9.5\times18\times16$ in. ($0.24\times0.46\times0.41$ m) aluminum enclosure intended to simulate a desktop computer tower or server chassis. A one inch wide rectangular obstruction has been mounted on the inlet side of the fan to create a time-invariant flow distortion. Four 1-1/8 in. (0.029 m) diameter Radio Shack loudspeakers surround the fan such that they are equidistant from each other and the distance between the center of the fan and the center of each loudspeaker is 0.06 m. The reference signal, which is required in the single reference, multiple output adaptive filtered-*x* controller, is supplied by locating a small infrared LED on the inlet of the fan and a matched phototransistor detector on the exhaust side of the fan, such that the blades passing between them creates a pseudosquare pulse train whose period corresponds to the BPF. Larson Davis Type-I microphones are mounted on the surface of the aluminum enclosure for error sensing and additional microphones are mounted on a rotatable semicircular measurement boom of radius 5 ft (1.52 m) for global performance monitoring. The 13 boom microphones are placed at equal angles along the boom and can therefore provide an estimate of the global mean-square pressure radiated by the fan at a given frequency.

IV. RESULTS AND DISCUSSION

The results of the theoretical analysis carried out have been utilized in conjunction with the above experimental apparatus to investigate the effect on the global attenuation of the fan by (1) a two-channel versus a four-channel ANC system; and (2) the placement of error microphones in ideal versus nonideal locations. Graphical results from both of the actuator configurations, meant to demonstrate typical reductions achieved with the error microphones located ideally, are first shown. These results are followed by numerical results and a discussion of the mean performance of each of the four configurations.

A. Graphical results

In order to determine the global mean-square pressure reductions achieved by a particular control configuration, a reference pressure measurement without ANC has been taken using the measurement boom. In Figs. 10 and 11, the directivity of the fan without control is represented by a wire frame mesh for each of the four targeted harmonics of the BPF, namely 370, 740, 1110, and 1480 Hz. The fan's radiation pattern becomes more directional and complex with increasing frequency, as the fan becomes less aeroacoustically compact.

Results are displayed in Fig. 10 for the first experimental configuration, a two-channel setup with the loudspeakers located on opposite sides of the fan and along the same axis, as described previously. The test channel count refers to the number of loudspeakers, error microphones, and control signals generated by the adaptive controller. A two-channel configuration will be often referred to hereafter as "2×2 control." For the trial shown in Fig. 10, the reductions in far-field mean-square pressure for the four harmonics of the BPF are 6.6, 9.8, 9.3, and 2.8 dB. The overall A-weighted tonal reduction [*dBAred* in Eq. (1)] is 7.5 dBA, which is calculated according to

$$\mathrm{dBAred}=10\log\left(\frac{\Sigma_{n=1}^{4}{p_{\mathrm{ANC\_OFF}}}^{2}(f_n)A^2(f_n)}{\Sigma_{n=1}^{4}{p_{\mathrm{ANC\_ON}}}^{2}(f_n)A^2(f_n)}\right), \qquad (1)$$

where $p^2(f_n)$ signifies the mean-square pressure for the nth harmonic of the BPF, represented by f_n, and A is the narrowband A-weighting function evaluated at the appropriate frequency. As with virtually all multichannel tests, the second harmonic, shown in Fig. 10(b), is reduced more than the fundamental, which is displayed in Fig. 10(a). In this case, the reduction of the third harmonic is also greater than that of the BPF. This may be attributed to the relatively poor response of the loudspeakers at 370 Hz.[24,25] Attenuation for the fourth harmonic, shown in Fig. 10(d), is much less significant than for the other three harmonics.

The global attenuations for a four-channel (4×4) test, again with the error microphones located ideally, are shown in Fig. 11. The mean-square pressure reductions for the four targeted frequencies are 9.5, 18.5, 14.4, and 7.6 dB, resulting in an overall tonal reduction of 11.4 dBA. Again, the attenuation of the second and third harmonics is greater than that of the fundamental, which may be seen by comparing Figs. 11(a)–11(c). The attenuation of the fourth harmonic, seen in

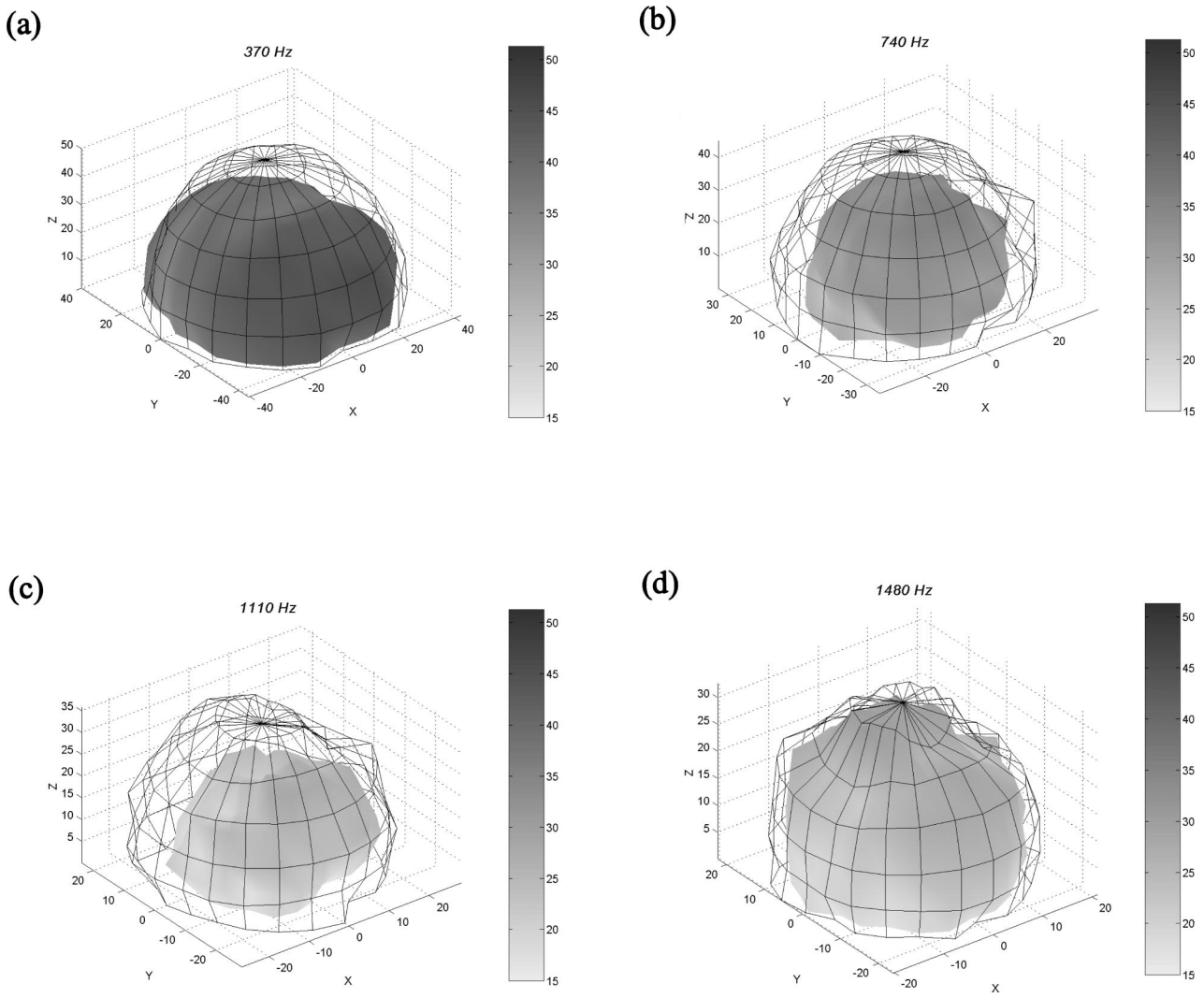


FIG. 10. Directivity of the BPF and harmonics without (mesh) and with (solid surface) two-channel active control, with the error microphones located in a theoretically ideal region. Both the radius from the origin and shading are proportional to the sound pressure level for a given angle.

Fig. 11(d), is only about 2 dB less than that of the fundamental, which again demonstrates the limitations imposed by the loudspeakers on the performance capabilities of the system at low frequencies.

B. Numerical results

The results displayed in the previous section are useful in understanding the capabilities of the multichannel ANC system for a single trial; however, little has been reported in the literature about the mean performance of active control systems. In Ref. 25, however, data are presented that show average mean-square pressure reductions of the tonal noise solely as a function of actuator configuration. Shown in Table I are the results of averaging the mean-square pressure attenuations for 2×2 and 4×4 control, for both the error microphones located along the theoretically ideal regions and also for nonideal placement. The results are obtained by selecting, for each configuration, the eight trials that yield the largest overall A-weighted tonal reductions. It is not possible to consider the mean performance of all trials because some nonideal error microphone locations quickly led to controller instability, and these divergent trials would need to be somehow included in the analysis to yield unbiased results. The results in Table I therefore serve as an average of the maximum mean-square pressure reductions yielded by the four configurations.

C. Discussion

The data in Table I demonstrate that, on the whole, the 4×4 setup provides a greater global attenuation of the fan tones than does the 2×2 configuration, corroborating the results of Ref. 25. In addition, placement of the error microphones along the theoretical nulls results in greater overall A-weighted tonal attenuations for both the two- and four-channel systems, although for the 2×2 arrangement, the nonideal locations yield greater mean-square pressure reductions for the third and fourth harmonics. While the tabular data serves to show quantitatively that greater reductions are possible with the error microphones located along the regions calculated to be theoretically ideal, some other qualitative observations are also worthwhile regarding system behavior as a function of error microphone placement.

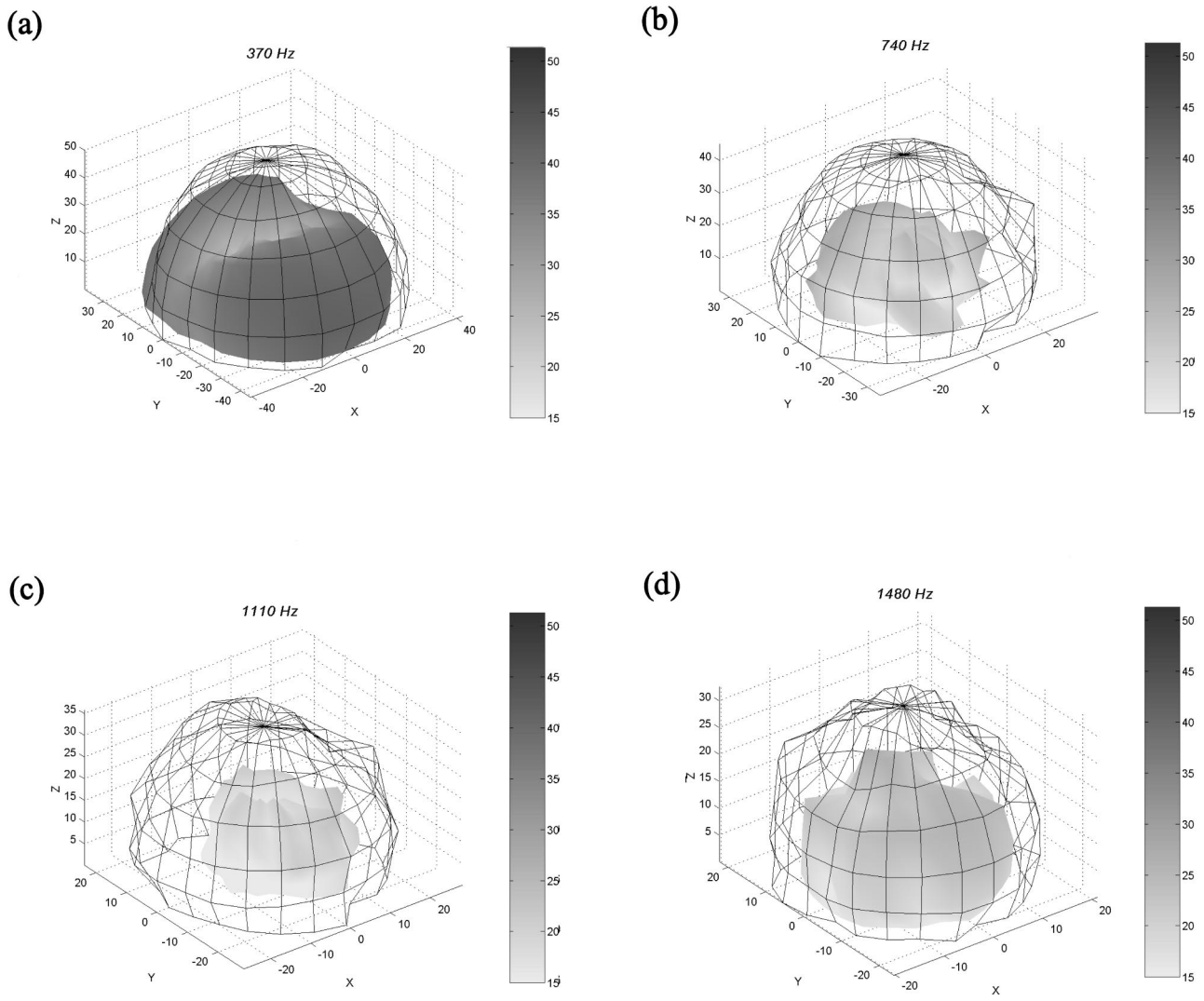


FIG. 11. Directivity of the BPF and harmonics without (mesh) and with (solid surface) four-channel active control, with the error microphones located in a theoretically ideal region. Both the radius from the origin and shading are proportional to the sound pressure level for a given angle.

Some comments with regards to controller stability, although qualitative in nature, help in fully describing the effect of error microphone placement on global tonal attenuations. Placement of the error microphones in locations determined to be theoretically ideal tends to result in control signals with smaller amplitudes and therefore a more stable controller; the adaptive controller very rarely diverges with the error microphones located ideally. However, nonideal (e.g., random) placement of the error microphones often produces large control signals and loudspeaker distortion, ultimately leading to positive feedback and divergence of the control system. This is especially true of the 4×4 system, whereas the 2×2 system tends to be more stable under loudspeaker distortion, but often yields net global increases in the mean-square pressure rather than attenuations. These results corroborate the findings of Baek and Elliott, which state that transducer positions resulting in low control effort are good choices when robust control is desired.[26]

TABLE I. Mean-square pressure reduction (MPR) and standard deviation (σ), in dB, averaged over the eight best trials for each type of configuration. ''Ideal'' signifies the error microphones are located along the theoretical nulls; ''nonideal'' means the error microphones are located somewhere off the nulls. Results for the first four harmonics of the BPF and the overall A-weighted tonal levels are shown.

	370 Hz		740 Hz		1110 Hz		1480 Hz		A-weighted Total	
All results in dB	MPR	σ	MPR	σ	MPR	σ	MPR	σ	MPR	σ
2×2 nonideal	5.3	2.5	10.9	5.0	6.9	3.9	4.2	2.5	6.5	2.7
2×2 ideal	8.4	2.5	11.9	3.1	5.9	2.7	3.1	3.0	8.0	1.2
4×4 nonideal	8.2	1.6	14.9	3.4	11.0	3.9	7.4	4.1	9.5	1.5
4×4 ideal	10.1	1.0	16.1	2.1	12.8	2.4	8.7	3.5	11.4	0.4

V. CONCLUSIONS

A theoretical modeling technique based on the previous work of Nelson *et al.*[16] has been applied to the multichannel active control of a small axial cooling fan. All sources have been represented as baffled ideal point sources and the control source strengths that minimize the overall radiated power have been found for various configurations. An examination of the radiated power as a function of nondimensional source separation distance shows that, under the restriction that all sources be coplanar and closely located, the amount of additional sound power reduction possible rapidly lessens as the number of symmetrically placed control sources is increased beyond three. This result suggests that a practical limit to the necessary number of symmetric control sources to maximize global attenuations from an acoustically compact coplanar noise source is three or four, depending on the frequency content of the primary source radiation. Moreover, for a given idealized point source configuration possessing relative source strengths such that the radiated power is minimized, an analysis of the radiated pressure in the plane of the sources has led to the discovery of ideal near-field error microphone locations.

These error microphone locations have been experimentally examined for two-channel and four-channel configurations and results describing the mean optimum performance of the active control system show that greater overall reductions may be achieved with the error microphones located in the theoretically ideal regions, rather than nonideal or random locations. Furthermore, ideal placement of the error microphones appears to lead to a more stable control environment, whereas random placement of the microphones often contributes to controller instability due to a large control effort and subsequent loudspeaker distortion. The results of this investigation show that significant global reductions of four harmonics of the BPF may be achieved upon utilization of an idealized model of the sources, and that the near-field placement of error microphones is plausible for global active control of the radiation from compact sources.

[1] G. H. Koopmann, D. J. Fox, and W. Neise, ''Active source cancellation of the blade tone fundamental and harmonics in centrifugal fans,'' J. Sound Vib. **126**, 209–220 (1988).

[2] R. H. Thomas, R. A. Burdisso, C. R. Fuller, and W. F. O'Brien, ''Active control of fan noise from a turbofan engine,'' AIAA J. **32**, 23–30 (1994).

[3] C. H. Gerhold, ''Active control of fan-generated tone noise,'' AIAA J. **37**, 17–22 (1997).

[4] R. Kraft, Z. Hu, S. Sommerfeldt, B. Walker, A. Hersh, H. Jo, M. Spencer, D. Hallman, C. Mitchell, and D. Sutliff, ''Development and demonstration of active noise control concepts,'' NASA CR-2000-210037, 2000.

[5] D. A. Quinlan, ''Application of active control to axial flow fans,'' Noise Control Eng. J. **101**, 95–101 (1992).

[6] G. C. Lauchle, J. R. MacGillivray, and D. C. Swanson, ''Active control of axial-flow fan noise,'' J. Acoust. Soc. Am. **101**, 341–349 (1997).

[7] M. Q. Wu, ''Active cancellation of small cooling fan noise from office equipment,'' *Proceedings of Inter-Noise 95*, Newport Beach, CA (Noise Control Foundation, Poughkeepsie, NY, 1995), pp. 525–528.

[8] X. Qiu, C. H. Hansen, and X. Li, ''A comparison of near-field acoustic error sensing strategies for the active control of harmonic free field sound radiation,'' J. Sound Vib. **215**, 81–103 (1998).

[9] C. H. Hansen and S. D. Snyder, *Active Control of Noise and Vibration* (E & FN SPON, London, 1997).

[10] 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).

[11] R. L. Clark and C. R. Fuller, ''Optimal placement of piezoelectric actuators and polyvinylidene fluoride error sensors in active structural acoustic control,'' J. Acoust. Soc. Am. **92**, 1521–1533 (1992).

[12] A. Berry, X. Qiu, and C. H. Hansen, ''Near-field sensing strategies for the active control of the sound radiated from a plate,'' J. Acoust. Soc. Am. **106**, 3394–3406 (1999).

[13] B.-T. Wang, ''Optimal placement of microphones and piezoelectric transducer actuators for far-field sound radiation control,'' J. Acoust. Soc. Am. **99**, 2975–2984 (1996).

[14] T. Martin and A. Roure, ''Active noise control of acoustic sources using spherical harmonics expansion and a genetic algorithm: Simulation and experiment,'' J. Sound Vib. **212**, 511–523 (1998).

[15] K. A. Cunefare and G. H. Koopmann, ''Global optimum active noise control: Surface and far-field effects,'' J. Acoust. Soc. Am. **90**, 365–373 (1991).

[16] P. A. Nelson, A. R. D. Curtis, S. J. Elliott, and A. J. Bullmore, ''The minimum power output of free field point sources and the active control of sound,'' J. Sound Vib. **116**, 397–414 (1987).

[17] T. Martin and A. Roure, ''Optimization of an active noise control system using spherical harmonics expansion of the primary field,'' J. Sound Vib. **201**, 577–593 (1997).

[18] J. S. Bolton, B. K. Gardner, and T. A. Beauvilain, ''Sound cancellation by the use of secondary multipoles,'' J. Acoust. Soc. Am. **98**, 2343–2362 (1995).

[19] T. A. Beauvilain, J. S. Bolton, and B. K. Gardner, ''Sound cancellation by the use of secondary multipoles: Experiments,'' J. Acoust. Soc. Am. **107**, 1189–1202 (2000).

[20] X. Qiu and C. H. Hansen, ''Secondary acoustic source types for active noise control in free field: Monopoles or multipoles?,'' J. Sound Vib. **232**, 1005–1009 (2000).

[21] W.-S. Chiu, G. C. Lauchle, and D. E. Thompson, ''Subsonic axial flow fan noise and unsteady rotor force,'' J. Acoust. Soc. Am. **85**, 641–647 (1989).

[22] K. B. Washburn and G. C. Lauchle, ''Inlet flow conditions and tonal sound radiation from a subsonic fan,'' Noise Control Eng. J. **31**, 101–110 (1988).

[23] P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic, London, 1992).

[24] K. L. Gee, ''Multi-channel active control of axial cooling fan noise,'' MS thesis, Brigham Young University, Provo, UT, 2002.

[25] K. L. Gee and S. D. Sommerfeldt, ''A compact active control implementation for axial cooling fan noise,'' Noise Control Eng. J., accepted for publication.

[26] K. H. Baek and S. J. Elliott, ''The effects of plant and disturbance uncertainties in active control systems on the placement of transducers,'' J. Sound Vib. **230**, 261–289 (2000).

Nonoccupational noise: exposures associated with routine activities

Richard Neitzel,[a] Noah Seixas, John Olson, William Daniell, and Bryan Goldman
University of Washington Department of Environmental and Occupational Health Sciences, Seattle, Washington 98195-7234

(Received 24 January 2003; revised 25 June 2003; accepted 11 August 2003)

Efforts to characterize nonoccupational noise exposures have focused primarily on infrequent, episodic events. Few studies have assessed noise levels resulting from routine daily activities. In the current study, 112 construction workers wore datalogging noise dosimeters and simultaneously completed activity logs during two phases of data collection. The 81 subjects monitored in phase 1 received logs listing numerous preselected occupational and nonoccupational activities, while the 31 subjects monitored in phase 2 used free-field logs and reported nonoccupational activities in greater detail. Nearly all of the 221 439 1-min intervals of nonoccupational L_{eq} level and activity reporting were below 70 dBA; only a small percentage exceeded 80 dBA. The primary contributor to nonoccupational noise exposure was traveling in a car or bus, while time at home contributed the least. One hundred seventy 24-h L_{eq} levels were computed from the 1-min noise level data. The percentage of phase 2 workday $L_{eq(24)}$ levels which exceeded 80 dBA was higher than that of the nonworkday levels. The mean $L_{eq(24)}$ level of phase 2 workdays was higher than that of nonworkdays, and the difference was statistically significant. Routine nonoccupational noise exposures contributed much less to total noise dose than occupational exposures in the subjects evaluated. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1615569]

PACS numbers: 43.50.Qp [DKW] Pages: 237–245

I. INTRODUCTION

Very few data exist describing routine daily noise exposures resulting from nonoccupational activities and their possible contribution to noise-induced hearing loss. It has been suggested that individuals with daily occupational noise exposures above 90 dBA (who represent less than 20% of US production workers[1,2]) have no additional risk of noise-induced hearing loss (NIHL) from their nonoccupational noise exposures, as these exposures are small when compared to their occupational exposure. However, the risk of NIHL resulting from nonoccupational noise alone probably increases in workers with occupational exposures below 90 dBA.[3] If this is so, then nonoccupational noise exposure may present a substantial fraction of the risk of hearing loss for US workers. However, such a conclusion requires additional information on the prevalence of high-level nonoccupational noise exposures; gathering this information was the goal of the current study.

Existing nonoccupational exposure level data primarily involve surveys of community noise levels or address infrequent noisy activities which may not be common to all individuals. Nonoccupational noise exposure data are usually collected either as weighted sound pressure levels or as exposure durations and/or frequencies, for one or more nonoccupational activities among a particular population. Rarely have both exposure level and activity frequency/duration data been collected in the same study, and in the few cases where they have been, only a small number of activities or exposure levels have been examined.[4] Exposures from activities assessed in the literature are typically high-intensity and the values reported are often various measures of peak or maximum sound pressure levels (SPLs) reached during a nonoccupational activity, rather than an average measure such as the equivalent continuous SPL (or L_{eq}), which is a more useful metric for regularly occurring, long-term exposures. For a review of previous nonoccupational noise exposure studies, the reader is referred to articles by Axelsson[5] and Clark.[6]

Due to inconsistencies in noise level documentation and a tendency to focus on the highest possible exposures during the noisiest nonoccupational activities, little is known about the exposure levels associated with typical daily activities. This paucity of data is due in part to the methodological challenges inherent in conducting a thorough nonoccupational noise exposure survey, and in part to historical technical limitations in measurement equipment, which until fairly recently have been too bulky or difficult to employ over extended periods of time off the job. Several studies have measured long-term noise exposures during occupational and nonoccupational activities;[7,8] however, these studies were limited by the small number of subjects evaluated and by a lack of detailed activity reporting, and have generally focused on 24-h L_{eq} levels ($L_{eq(24)}$) only, and not activity-specific levels. Other studies have described sound pressure level ranges associated with common nonoccupational equipment,[6] recreational activities,[5] and public transit and general environmental noise,[9] but these data are presented as a catalog of potential noise levels rather than a comprehensive survey of exposures in a defined population.

Construction workers have been shown to have high daily occupational noise levels,[10,11] and are therefore at risk of NIHL from exposures at work. High nonoccupational noise exposures have the potential to increase this risk and

[a]Electronic mail: rneitzel@u.washington.edu

affect individuals' cumulative noise dose if they are encountered frequently enough to substantially contribute to overall noise dose. The current study used a combination of datalogging dosimetry and activity logs to assess the levels associated with routine nonoccupational noise exposures encountered by a cohort of construction workers.

II. METHODS

A cohort of apprentice construction workers from a variety of trades in the Puget Sound area of the state of Washington was assembled from 1999 to 2000 to participate in a 5-year longitudinal study of noise exposure and hearing loss. As part of this study, nonoccupational noise exposures were assessed through a combination of dosimetry measurements and activity logs and questionnaires. Only the dosimetry measurements and activity logs will be reviewed here; the results of the questionnaire concerning the frequency with which noisy activities outside work are conducted will be discussed by the authors in a future publication. Dosimetry measurements and simultaneous activity log data were collected in two different phases: phase 1 involved 40 hours of consecutive noise monitoring and reporting on individual subjects, and assessed occupational and nonoccupational noise exposures, while phase 2 involved 96 consecutive hours of measurement and reporting and focused primarily on nonoccupational exposures. All subjects attended a brief overview session describing the study; interested volunteers signed an IRB-approved informed consent form prior to enrollment in the study. A small monetary incentive was offered to increase participation rates. Dosimetry and activity log data were combined to allow for noise exposure analysis by activity and by 24-h exposure period. Eighty-one workers from eight trades participated in the phase 1 exposure assessment: 6 masonry restoration workers, 11 carpenters, 10 cement masons, 17 electricians, 9 insulation workers, 10 ironworkers, 9 sheet metal workers, and 9 tilesetters. Phase 1 subjects were monitored on workdays only, and worked at construction sites. Thirty-one workers from six trades participated in phase 2: 8 bricklayers, 4 masonry restoration workers, 4 carpenters, 6 ironworkers, 6 sheet metal workers, and 3 tilesetters. Phase 2 subjects were monitored on both workdays and nonworkdays, and their workdays consisted of work at construction sites or training at apprenticeship training centers.

A. Noise measurements in phases 1 and 2

Noise exposure levels during both phases of data collection were measured using Quest Q-300 Type 2 datalogging noise dosimeters. These units were configured to measure the equivalent continuous sound level, or L_{eq}, using a slight modification of the settings specified by the 1998 NIOSH Recommended Exposure Limit (REL) (85 dBA criterion level, 3 dB exchange rate, 70–140 dBA measurement range, and fast response).[1] The only settings that differed from the REL were the threshold level, which was set at 0 dBA (versus 80 dBA for the REL) and, in phase 2, the measurement range of the dosimeters. During phase 1 of data collection dosimeters were set to a measurement range of 70–140 dBA, as the data being collected were from both occupational and nonoccupational exposures, and occupational construction noise levels were expected to be high. However, during phase 2 of data collection, the dosimeters were set to a measurement range of 40–110 dBA, as the data of interest were nonoccupational, and these exposure levels were expected to be lower than those encountered occupationally.

The dosimeters yielded an L_{eq} level for each 1-minute interval measured. These 1-min interval noise levels, in combination with subjects' reported activities, are the basis of the activity-specific exposure level data presented here. Phase 1 1-min intervals in which the sound pressure level was below 70 dBA were recorded by the dosimeters as 69.9 dBA, while phase 2 1-min intervals in which the sound pressure level was below 40 dBA were recorded as 39.9 dBA. The presence of these limit-of-detection values means that the data discussed here overestimate exposure levels for activities with low associated sound pressure levels. The amount of overestimation is larger for phase 1 data than for phase 2 due to the higher limit-of-detection level in phase 1. The difference in the limit-of-detection level prevents the combination of phase 1 and phase 2 data, and the mean and percentile noise level data are therefore presented and discussed separately here. However, both measurement ranges used measure noise levels between 70 and 110 dBA accurately; therefore, exceedance percentages within this range are presented for the phase 1 and 2 1-minute average levels separately and in combination.

The maximum operating and datalogging time of the dosimeters was limited both by battery life and memory capacity. In order to accommodate these limitations and prevent loss of data, no dosimeters were run for longer than 48 consecutive hours.

During phase 1 of data collection, subjects were instructed to turn on their dosimeters at the time they awoke on the first workday that they had the dosimeter. The dosimeters were programmed to log data for the next 40 consecutive hours (a period intended to cover two full workshifts and the intervening nonwork period), and to stop datalogging at the end of that period. At the end of the monitored period units were retrieved from subjects by research staff. During phase 2, subjects were issued two dosimeters simultaneously, and were instructed to turn on the first dosimeter at the time they awoke on the first Thursday that they had the dosimeter. The first dosimeter was worn for a period of 48 h, at which time the instrument stopped datalogging. Subjects were instructed to doff the first dosimeter and don the second dosimeter at the 48-h mark, and then to wear the second dosimeter for an additional 48 h, at which point the second dosimeter stopped datalogging. Monitoring therefore occurred over a total of 96 h (Thursday morning to Monday morning). This technique was used to incorporate weekend time into each sample and maximize the measured duration of nonoccupational activities.

Dosimeters were worn on subjects' belts or in their pockets, microphones (with windscreens) were attached to the subject's collar or lapel, and subjects were advised not to cover the microphone with clothing or to remove the dosimeter once monitoring had begun unless they were sleeping, bathing, or performing rigorous activities in which the do-

simeter would interfere. Subjects were instructed to keep the dosimeters as close as possible to their person when it was necessary to remove them, so the instruments could continue to measure representative exposure levels. All dosimeters were calibrated pre- and post-monitoring. Dosimeters were downloaded directly into a PC using QuestSuite™ software.

B. Activity reporting phase 1 (40 h)

Phase 1 of data collection included both nonoccupational and occupational activities. A single activity log was used to report both types of activities. Start and end times were recorded on a continuous timeline. Because of the condensed nature of the logs (two sheets covering 24 h each), the actual times recorded and extracted from the timelines were in approximately 15-min intervals. In order to keep the log to a manageable size and complexity, only a limited number of preselected activities in both the occupational and nonoccupational categories could be provided, although subjects were able to report an additional two occupational and two nonoccupational activities not already listed. The six preselected nonoccupational activities provided on each log were bar/restaurant/shopping/theatre, home, listen to music/watch TV, travel in a car/bus, yardwork, and other. These activities were chosen because they were expected to represent the majority of any individual subject's time, and covered a broad range of activity types. Subjects were provided with the appropriate trade-specific activity log when they were issued a dosimeter. Each trade-specific log included a list of up to eight work activities typically encountered in their trade. The use of this type of activity log in occupational settings has been validated previously over the course of a single day,[10,11] and over a longer period.[12] Similar approaches have proven successful for studying other exposures, such as environmental pollutants.[13]

C. Activity reporting phase 2 (96 h)

Phase 2 of the data collection focused solely on nonoccupational activities, though the monitored periods did encompass periods of occupational activity. As with phase 1, subjects used a single activity log to record start and stop times for both types of activities with a continuous 24-h timeline per sheet (four sheets total per log) and a similar time resolution of approximately 15 min. However, the log used a free-field reporting format for identifying activities. The log allowed subjects to report their location, the number of people in their immediate area, and the perceived loudness of activities and noise sources. The log also included a free field in which subjects were instructed to document any noise sources in their area and to provide a brief description of their activities. Subjects were instructed to report their activities and sound environment conditions on the activity log, with the exception of time spent in an occupational setting, in which case they were instructed to simply note the times at which they began and ended work.

D. Data coding

One-minute interval L_{eq} noise level data were exported from Questsuite™ as text files. Activity log data were entered into an MS Access database. The noise level data were then merged into the database, based on subject ID, date, and time. Nearly 100% of phase 1 nonoccupational activity log data consisted of the six precoded nonoccupational activities from the predetermined lists on each trade-specific log; the few data provided for additional activities not listed on the cards were coded as ''Other.''

The more detailed free field activity log data collected in phase 2 required additional coding procedures. Activities and equipment were defined during data coding based on the free-field information provided. Free-field nonoccupational activities were coded into 37 categories with similar expected noise levels; these categories were then collapsed into 11 activity groups, and then were collapsed further into the six activities used in phase 1. Occupational activities were generically coded as work time.

Some 1-min intervals in phases 1 and 2 had more than one reported activity during certain 1-min intervals. To account for the occurrence of simultaneous activities and insure that total minutes in a monitored day equaled 24 h, a weighting system was utilized, with minutes in which a single activity occurred receiving a weight of 1, and minutes of data in which more than one activity was reported receiving a weight equal to (1/number of simultaneous activities). For example, each portion of a minute in which three simultaneous activities were reported would be assigned a weight of 0.33. Each portion of all minutes with a weight less than 1 was assigned the measured sound pressure level corresponding to that minute.[15] Once all nonoccupational activity data were entered and the 1-min interval weighting was implemented, the MS Access database tables were then imported into SAS version 8.2 for statistical analysis.

E. Data analysis

Descriptive statistics were generated by activity for phase 1 1-min interval L_{eq} noise levels. In phase 2, analyses were also conducted by individual and grouped activity, location, number of people in the immediate area, and perceived loudness. Phase 1 and 2 work time data were removed from the database for these descriptive 1-min L_{eq} level analyses. One-minute activity log data and their corresponding 1-min L_{eq} level data were aggregated by subject, and activity-specific L_{eq} exposure levels ($L_{eq_{Aij}}$) for each subject were calculated using Eq. (1):

$$L_{eq_{Aij}} = 10 \log_{10} \frac{1}{n_{ij}} \sum_{k=1}^{N} 10^{L_{eq_{ijk}}/10}, \tag{1}$$

where n_{ij} is the number of minutes that subject j conducted activity i, and $L_{eq_{ijk}}$ is the SPL recorded for subject j and activity i in minute k. Mean L_{eqA} levels over all subjects reporting the activity were then calculated from these activity-specific subject levels. Exceedance percentages were also developed from the 1-min interval L_{eq} noise level data.

Because phase 1 noise levels below 70 dBA were recorded as 69.9 dBA, and phase 2 noise levels below 40 dBA were recorded as 39.9 dBA, the calculated mean noise levels are biased, though the bias in the phase 2 data is minimal. This inherent bias makes interpretation of the mean levels

TABLE I. Summary of phase 1 and 2 noise level data.

	Phase 1 (8 trades, 40-h monitoring period)			Phase 2 (6 trades, 96-h monitoring period)		
	Work	Nonwork	Total	Work	Nonwork	Total
All 1-min noise levels						
No. of minutes	88 456	105 944	194 400	30 179	148 381	178 560
No. of subject-days	159	162	162	58	124	124
No. of subjects			81			31
1-min noise levels with reported activities						
No. of minutes	79 681	92 973	172 654	22 701	128 466	151 167
No. of subject-days	154	156	156	56	118	118
No. of subjects			81			31
1-minute noise levels used for 24-h L_{eq}'s						
No. of minutes	43 337	73 343	116 640	22 343	106 123	128 466
No. of subject-days	81	81	81	53	36	89
No. of subjects			81			31

somewhat difficult; for example, the mean noise levels calculated are likely somewhat higher than actual values, particularly in phase 1 activity data where few minutes have levels exceeding 70 dBA. To avoid this bias, frequency tables with 10 dBA increments were developed in addition to mean levels.

Individual 24-h L_{eq} ($L_{eq(24)i}$) levels were calculated to assess the total noise resulting from the combination of occupational and nonoccupational noise exposure in the measured subjects. $L_{eq(24)i}$ levels were generated by identifying subject-specific sequences of consecutive 1-min intervals that were approximately 24 h (1350 to 1440 min) in length, and calculating the overall 24-h mean of the 1-min L_{eq} sound pressure levels measured during each run, using the methods described above and Eq. (2):

$$L_{eq(24)i} = 10 \log_{10} \frac{1}{n_i} \sum 10^{L_{eq_{ik}}/10}, \tag{2}$$

where n_i is the duration of the measured run (1350 to 1440 min for an $L_{eq(24)}$) for individual i and $L_{eq_{ik}}$ is the noise level for individual i during 1-min noise interval k. Mean $L_{eq(24)}$ and exceedance percentages across the study population were developed using the $L_{eq(24)i}$ values. One-minute intervals associated with occupational activity were included in the $L_{eq(24)i}$ noise level calculations, and $L_{eq(24)}$ levels were stratified by workdays (24-h periods that included work time) and nonwork days (no work time reported). Workday $L_{eq(24)}$ levels were further stratified into work time and nonwork time, and L_{eq} noise levels were developed for both of these components using Eq. (2). Exceedance percentages were also developed for the $L_{eq(24)}$ data. Nonworkday $L_{eq(24)}$ levels could not be developed from the phase 1 data, as all 24-h periods included work time.

III. RESULTS

A summary of the phase 1 and 2 data is shown in Table I. Mean subject age was 29.6 ± 6.2 years, and 85% of subjects were male. One hundred twelve subjects completed 81 phase 1 and 31 phase 2 samples, representing 194 400 and 178 560 min of monitoring, respectively. Nine instrument failures occurred, for a failure rate of 8%. Although subjects were instructed to always record in their activity log all activities in which they were involved, about 11% of phase 1 and 15% of phase 2 1-min interval dosimetry data had no recorded activity. It was impossible to infer the activities of subjects during these unreported periods, so these 1-min intervals were removed from the dataset. This reduced the phase 1 data to 172 564 reported minutes and the phase 2 data to 151 167 reported minutes. A subset of these reported 1-min intervals were from nonoccupational activities (221 439 minutes total). Reported phase 1 data represented 156 subject-days of monitoring, and reported phase 2 data represented 118 subject-days.

For $L_{eq(24)}$ analysis, only blocks of 1-min intervals between 1350 and 1440 min in length were utilized; however, work time was included in this analysis. The time restriction was lowered to 1350 min to increase sample size, as a number of blocks of useable data did not cover a full 1440 continuous minutes. The time block restriction and inclusion of work time resulted in a phase 1 dataset of 81 24-h periods which included work time (116 640 1-min intervals), and a phase 2 dataset of 53 24-h periods which included work time and 36 which included no work time (89 total phase 2 24-h periods with 128 466 1-min intervals).

Noise level data for the six routine nonoccupational activities assessed during phases 1 and 2 of data collection are presented in Table II. The mean phase 1 levels for the six activities are all close to 70 dBA, the limit of detection. Seventy-three percent of all phase 1 1-min interval L_{eq} levels were under 70 dBA. Two of the activities, travel in a car or bus and bar/restaurant/shopping/theatre, had higher phase 1 mean levels and exceedance percentages than the other activities. Eighty percent of all 1-min interval noise level data from phase 2 were below 70 dBA. The phase 2 L_{eqA} mean level was 74.0 dBA. The activity with the highest associated L_{eqA} noise level in phase 2 was yardwork, and 18% of all phase 2 minutes associated with travel in a car or bus ex-

TABLE II. Phase 1 and 2 nonoccupational noise levels by routine activity.

Activity	Data phase	Total min	% Total	Mean L_{eqA} level (dBA)[a,b]	1-min L_{eq} level percentages			
					%⩽70 dBA	70<%⩽80 dBA	80<%⩽90 dBA	%>90 dBA
Bar/restaurant/ shopping/theatre	1	3046	3	74.9	39	41	18	3
	2	1220	1	71.4	67	28	5	0
	Combined	4266	2		47	37	14	2
Home	1	52 123	56	70.6	88	10	2	0
	2	57 140	45	66.8	91	7	2	0
	Combined	109 263	49		90	8	2	0.3
Listen to music	1	10 145	11	72.6	65	23	10	2
	2	42 035	33	73.1	79	15	5	1
	Combined	52 180	24		79	15	5	1
Other	1	11 262	12	71.4	78	15	5	1
	2	19 906	16	75.8	70	21	8	1
	Combined	31 168	14		75	17	7	1
Travel in a car/bus	1	14 872	16	76.1	32	41	22	5
	2	7234	6	78.1	52	31	15	3
	Combined	22 106	10		39	37	20	4
Yardwork	1	1525	2	73.5	54	31	13	2
	2	930	1	78.5	64	30	4	2
	Combined	2455	1		58	30	10	2
Total	**1**	**92 973**	**100**	**74.9**	**73**	**18**	**7**	**2**
	2	**128 466**	**100**	**74.0**	**80**	**14**	**5**	**1**
	Combined	**221 439**	**100**		**79**	**15**	**5**	**1**

[a]Positively biased because phase 1 L_{eq}<70 dBA were recorded as 69.9 dBA.
[b]May be slightly biased because phase 2 L_{eq}<40 dBA were recorded as 39.9 dBA.

ceeded 80 dBA. Overall, 79% of phase 1 and 2 nonoccupational minutes were below 70 dBA, and only 1% exceeded 90 dBA. Home was associated with the lowest phase 1 and 2 mean levels and highest percentage (90%) of phase 1 and 2 nonoccupational reported minutes below 70 dBA, while travel in a car or bus was associated with the highest percentage (24%) of reported minutes above 80 dBA.

The 170 phase 1 workday and phase 2 workday and nonworkday 24-h L_{eq} levels are shown in Table III. The mean $L_{eq(24)i}$ duration was 1432 min. The $L_{eq(24)}$ data were grouped by the presence or absence of work activity during the measured period, and, for workday $L_{eq(24)}$ levels, by the work and nonwork time reported within the 24-h period. Roughly one-third of phase 1 and 2 workday time, on average, was spent at work. The effect of the phase 1 70 dBA limit-of-detection is apparent in the exceedance percentages; 99% of phase 1 workday $L_{eq(24)}$ levels were above 70 dBA, and 34% were above 85 dBA. The mean phase 2 workday $L_{eq(24)}$ was 75.0 dBA, with 7% of workday $L_{eq(24)}$ samples above 85 dBA, while the mean phase 2 nonworkday $L_{eq(24)}$

TABLE III. Phase 1 and 2 $L_{eq(24)}$ and occupational and nonoccupational time L_{eq} noise levels.

Date phase	Category	No. subject-days/mean % time of days	No. minutes	Mean $L_{eq(24)}$ noise level (dBA)[a]	L_{eq} level exceedance percentages				
					%>70 dBA	%>75 dBA	%>80 dBA	%>85 dBA	%>90 dBA
1	Workday $L_{eq(24)}$[a]	81	116 640	83.0	99	95	72	34	8
	Within workday: Work time L_{eq}	37%	43 337	85.0	98	96	79	50	22
	Nonwork time L_{eq}	63%	73 343	77.3	99	60	21	9	1
2	Nonworkday $L_{eq(24)}$[b]	36	52 107	72.0	69	36	19	3	2
	Workday $L_{eq(24)}$[b]	53	76 359	75.0	81	53	27	7	3
	Within-workday: Work time L_{eq}	29%	22 343	79.3	89	72	47	25	9
	Nonwork time L_{eq}	71%	54 016	72.7	70	34	15	2	<1

[a]Positively biased because phase 1 L_{eq}<70 dBA were recorded as 69.9 dBA.
[b]May be slightly biased because phase 2 L_{eq}<40 dBA were recorded as 39.9 dBA.

TABLE IV. Phase 2 nonoccupational noise levels by activity.

Activity group	Activity	Total minutes	% total	Mean L_{eqA} level (dBA)[a]	1-min L_{eq} level percentages			
					%⩽70 dBA	70<%⩽80 dBA	80<%⩽90 dBA	%>90 dBA
Home labor		945	0.7	91.4	65	29	4	2
	Cutting wood	30	0.0	82.1	63	30	0	7
	Garage work	315	0.2	75.9	65	29	9	0
	Mechanic work	60	0.0	68.7	83	17	0	0
	Repairs	90	0.1	62.9	94	6	0	0
	Washing car	210	0.2	72.0	52	48	1	0
	Yard work	240	0.2	97.2	63	27	4	6
Housekeeping		2118	1.6	74.6	68	26	6	0
	Chores	145	0.1	74.7	22	74	4	0
	Cooking	757	0.6	73.0	73	22	4	0
	Cleaning	240	0.2	70.2	78	21	1	0
	Walking dog	87	0.1	74.0	87	9	2	1
	Housekeeping	470	0.4	76.2	63	32	5	1
	Laundry	420	0.3	76.1	70	17	13	0
Lounging		4013	3.1	73.3	84	12	4	0
	Lounging	868	0.7	71.3	86	10	3	0
	Reading	775	0.6	70.3	95	5	0	0
	Talking	2331	1.8	74.5	81	15	5	0
	Smoking	39	0.0	76.2	15	74	10	0
Entertainment		46 969	37	77.0	79	79	15	5
	Boardgame	480	0.4	65.2	95	4	1	0
	Computer use	480	0.4	68.8	89	9	2	0
	Phone	205	0.2	75.8	47	41	12	0
	Radio	13 404	10	80.2	60	25	13	3
	Television	31 103	24	74.7	87	11	2	0
	Videogame (home)	1297	1	72.7	82	16	2	0
	Movie theatre	330	0.3	71.9	53	45	2	0
	Playing music	16	0.0	79.5	31	38	31	0
Sporting event		420	0.3	73.6	82	82	14	3
Shooting		150	0.1	81.7	5	52	42	1
Sports		540	0.4	66.5	88	11	1	0
	Fishing	60	0.0	72.7	37	62	2	0
	Golf	420	0.3	64.8	94	5	1	0
	Sledding	30	0.0	46.4	100	0	0	0
	Cross-country skiing	30	0.0	45.2	100	0	0	0
Shopping		560	0.4	75.3	70	70	23	7
Life functions		4683	3.7	75.0	70	70	20	10
	Bathing	702	0.8	72.2	81	15	4	0
	Drying hair	43	0.0	79.1	49	49	0	2
	Eating	3908	3.0	75.3	73	20	6	1
	Haircut	30	0.0	68.8	70	30	0	0
Sleeping		46 413	36	48.8	96	3	1	0
Other		21 308	17	78.8	71	20	8	1
Total		**128 466**	**100.0**	**74.0**	**80**	**14**	**5**	**1**

[a]May be slightly biased because phase 2 $L_{eq}<40$ dBA were recorded as 39.9 dBA.

was 72.0 dBA, with 3% of nonworkday $L_{eq(24)}$ samples above 85 dBA.The phase 2 nonwork periods during workdays had a mean L_{eq} level of 72.7 dBA, almost identical to the mean $L_{eq(24)}$ level for nonwork days. The difference between the mean phase 1 L_{eq} levels associated with work and nonwork time was statistically significant (t-test, $p<0.0001$), and the percent of within-workday L_{eqi} levels exceeding 75, 80, 85, and 90 dBA was substantially higher for work time than for nonwork time. The difference between the mean phase 2 $L_{eq(24)}$ level of workdays and nonworkdays was also statistically significant (t-test, $p<0.0001$), as was the difference between phase 2 within-workday work and nonwork time.

Table IV shows the distribution of noise level data for all 11 activity groups and 37 individual activities from phase 2 of data collection. The limit-of-detection bias in the phase

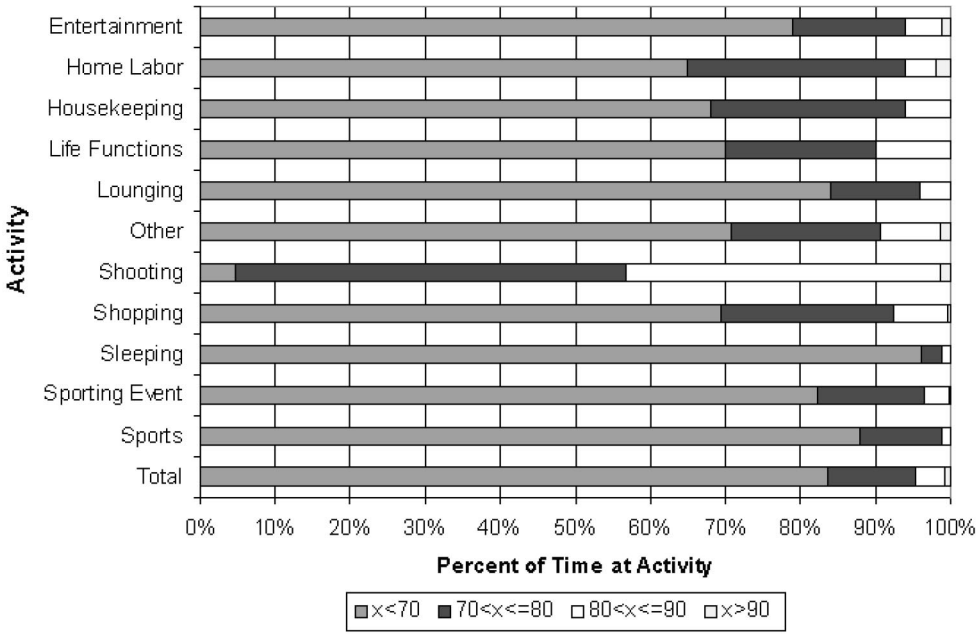


FIG. 1. Distribution of phase 2 1-min L_{eq} levels by self-reported nonoccupational activities.

2 L_{eqA} data is essentially negligible, as noise levels below 40 dBA are uncommon in an urban setting.[16] Roughly 43% of the 150 1-min intervals during which shooting was reported exceeded 80 dBA—the highest for any individual activity, followed by playing music (31%). The individual activities with the highest associated L_{eqA} noise levels were yardwork and shooting, and the activity group with the highest associated L_{eqA} noise level was home labor; however, few data were collected during these activities, so these findings may be somewhat unstable. Figure 1 illustrates the distribution of 1-min interval L_{eq} sound pressure levels across all reported phase 2 activities. At least 65% of reported 1-min intervals in all activities other than shooting had L_{eq} levels below 70 dBA, and 90% of all data for activities other than shooting were below 80 dBA.

Table V shows the perceived sound levels reported by subjects in phase 2 and the associated dosimetry noise level data. Subjectively assessed exposure levels corresponded well with the measured noise levels. Only a very small percentage of 1-min intervals reported as quiet exceeded 70 dBA, while a higher proportion of minutes with a perceived moderate sound level, and a higher-still proportion of minutes with a perceived sound level of loud exceeded 70 dBA. The mean L_{eq} levels associated with the categorical perceived noise level responses also rose appropriately, with the perceived intermittent loud exposure (79.6 dBA) falling between moderate (74.1 dBA) and loud (83.2 dBA).

IV. DISCUSSION

The noise levels measured in the current study suggest that the majority of time (70% or more) spent in nonoccupational activities is associated with an equivalent continuous exposure level below 70 dBA. The mean L_{eqA} level across all phase 2 nonoccupational activities measured here was 74.0 dBA, and in both phases of data collection less than 10% of total nonoccupational time was spent above 80 dBA. The activity associated with the largest percentage of minutes spent above 70 dBA (and the highest mean L_{eqA} noise level in phase 1) was traveling in a car or bus, while the activity associated with the lowest mean L_{eqA} noise level and smallest percentage of minutes spent above 70 dBA was time at home, including sleep time. Certain regular activities, such as traveling in a car or bus, are more likely to involve exposure to noise levels greater than 90 dBA than are other activities, though these levels are only associated with slightly

TABLE V. Phase 2 nonoccupational noise level distribution by reported location and sound environment.

Perceived sound level	Total minutes	% Category	Mean L_{eqA} level[a]	1-min L_{eq} level percentages			
				%≤70 dBA	70<%≤80 dBA	80<%≤90 dBA	%>90 dBA
Quiet	77 315	60	66.7	93	5	2	0
Moderate	39 308	31	74.1	70	22	7	1
Intermittent loud	300	0.2	79.6	57	33	9	0
Loud	4478	4	83.2	46	30	17	7
None reported	7065	6	76.8	78	15	7	1
Total	**128 466**	**100.0**	**74.0**	**80**	**14**	**5**	**1**

[a]May be slightly biased because phase 2 L_{eq}<40 dBA were recorded as 39.9 dBA.

over 11% of the total time reported by the cohort. Listening to music, an activity often associated with a high risk of NIHL, was not associated with particularly high average exposure levels or exceedance percentages in this study. High-level exposure to music has been shown to produce temporary[16] and permanent[17] threshold shifts, and can certainly represent a hearing loss risk with exposure of sufficient magnitude and duration. However, the exposures measured in the current study suggest that these high-level, long duration exposures are rare, a conclusion which has also been drawn from other research.[5,6]

Because some phase 2 workdays measured in this study included training time in apprenticeship training centers, the phase 2 workday L_{eq} levels (mean 79.3 dBA), were lower than the phase 1 levels (mean 85.0 dBA) which were measured on construction sites. The difference in work levels between phases 1 and 2, combined with the higher limit of detection used in the phase 1 measurements, resulted in a difference of 8 dB between the phase 1 and 2 workday $L_{eq(24)}$ levels (83.0 and 75.0 dBA, respectively). For further information on full-shift exposure levels in construction workers, the reader is referred Refs. 10 and 11.

The subjects measured in phase 2 of the current study had a mean workday $L_{eq(24)}$ of 75.0 dBA (including time spent doing construction work and apprenticeship training), and a mean nonworkday $L_{eq(24)}$ of 72.0 dBA. Eighty-one percent of phase 2 workday $L_{eq(24)}$ measurements exceeded 70 dBA, and 27% exceeded 80 dBA (energetically equivalent to an 8-h L_{eq} of 85 dBA), while 69% and 19% of nonworkday $L_{eq(24)}$ measurements were above 70 and 80 dBA, respectively. On average, the percentage of $L_{eq(24)}$ measurements above each exceedance level from 75 to 90 dBA was roughly twice that of nonworkday $L_{eq(24)}$ measurements. Work-related exposure levels were statistically significantly higher than the nonworkday levels. Noise exposure at work accounted for only one-third of workday exposure time on average, but contributed the majority of the noise exposure in both phase 1 and 2 workday $L_{eq(24)}$ measurements.

The nonoccupational exposure levels measured in phase 2 of the current study are slightly higher than the limited data available in the existing literature on routine nonoccupational exposures.[7–9,18] Conversely, listening to music, an activity which has been associated by some authors with high exposure levels[3,5] and an elevated risk of NIHL,[17] was not associated with high mean exposure levels or high exceedance percentages in this study. Although a high percentage of phase 2 $L_{eq(24)}$ exposures were above the EPA and WHO recommended 24-h exposure limit of 70 dBA, and are therefore considered potentially damaging to hearing, this recommended exposure limit has an inherent safety margin designed to protect the most sensitive individuals, and is overprotective for most individuals.[8] At these lower noise levels, nonauditory effects may be more important than risk of hearing loss.[19]

The current study confirms the suggestion in previous research that certain activities contribute the majority of an individual's total noise dose. Johnson and Farina,[7] the first researchers to consider and successfully conduct long-term noise dosimetry measurements incorporating both occupational and nonoccupational exposures, monitored the noise exposure of a single individual for 31 consecutive days. Exposure at parties and nightclubs (85.4 dBA L_{eq}) and during bowling (85.5 dBA L_{eq}) and automotive shop hobby work (90.5 dBA L_{eq}) accounted for less than 10% of the subject's total time, but contributed 69% of the total noise dose. The average noise level for the entire 31-day period was 76 dBA, with 24-h equivalent continuous noise levels ($L_{eq(24)}$'s) ranging from 59 to 83 dBA.

Schori and McGatha[8] found that, in 2 of 50 subjects (median age 22.4 years) evaluated, the majority of the measured 7-consecutive-day ($L_{eq(week)}$) noise exposure came from activities that accounted for less than 5% of the measured subjects' time (i.e., using a jig saw and attending a rock concert). As with the current study, a number of the highest L_{eq} sound levels measured by Schori and McGatha were from subjects riding in cars. These subjects spent an average of 2.3% of their total time in the car, but this exposure on average accounted for nearly 10% of their total noise exposure. The authors noted that only 1 of the 50 weekly levels (which generally ranged from 73 to 76 dBA) measured would approach the weekly exposure of a worker with an average occupational exposure of 90 dBA.

The use of activity logs to collect nonoccupational activity information requires subjects to self-report their activities. Although the accuracy and validity of this approach has been demonstrated in previous occupational noise studies,[10,11,13] validating the accuracy of nonoccupational activity reporting would be invasive and prohibitively time-consuming. Reporting biases and errors may have occurred if the logs were not completed on an ongoing basis, but were rather completed retrospectively at one or more times during the monitored period. Despite the detailed activity log reporting instructions given to subjects and attempts to clarify unclear entries with subjects upon log receipt, some activity logs contained data that could not be interpreted or entered into the database. The use of a predetermined activity list in phase 1 of data collection may have resulted in misclassification of data. Conversely, the use of a free field activity section in phase 2 of data collection introduced a great deal of variability in reported activities. Although every effort was made to insure consistency during the coding of these free-field-reported activities, misclassification certainly occurred. The use of multi-day activity logs prohibited the use of a minute-by-minute activity reporting timeline, and the continuous timeline utilized introduces additional error into the data collection, although the error is presumably random and results in both under- and overreporting of activity duration. Very few data were collected on certain activities like shooting and yardwork. There are additional limitations of the noise level data associated with shooting. Modern dosimeters (such as the Q-300s used in this study) utilize microphones which generally cannot measure peak sound pressure levels above 140–150 dB, well below the peak levels associated with some firearms.[20] The use of A-weighted equivalent average levels may underestimate the hazards associated with impulsive noise such as that of shooting.[21] In addition, high level impulse sounds produce hearing damage through a different mechanism (instantaneous acoustic trauma produc-

ing direct mechanical damage) than do continuous, lower-level sounds (which produce sensorineural damage with long-term exposure).[22] Lastly, the high level at which the phase 1 data were censored (70 dBA) resulted in overestimates of actual phase 1 exposure levels, and made the interpretation and use of this data problematic. The phase 2 data, however, are far more reliable and accurate, and can be used for further modeling of nonoccupational exposures.

V. CONCLUSION

Sources and levels of occupational noise exposure in the construction industry have been relatively well characterized.[10,11,23–28] However, nonoccupational noise exposures have not been adequately explored in this industrial sector or others. This study assessed a number of activities reported by construction workers and the noise levels associated with these activities. The data presented here suggest that, for most individuals in the construction industry, the predominance of noise exposure stems from occupational, rather than routine nonoccupational, activities, and that most noise exposure during routine non-occupational activities is below 70 dBA. Nevertheless, routine nonoccupational activities such as traveling in a car or bus can occasionally involve exposure to high noise levels. Individuals with long-duration exposures to these noisy routine activities could receive substantial noise exposures approaching, or on par with, their occupational exposures. Attempts to minimize the relatively infrequent high-level nonoccupational noise exposures should receive additional attention.

[1] National Institute for Occupational Safety and Health, Criteria for a Recommended Standard: Occupational Noise Exposure, Report No. DHHS (NIOSH) 98-126 (1998).

[2] J. R. Franks, ''Number of workers exposed to occupational noise.'' Seminars in hearing **9**(4), 287–297 (1988).

[3] Research Medical Research Council Institute of Hearing, ''Damage to hearing arising from leisure noise,'' Br. J. Audiol. **20**(2), 157–164 (1986).

[4] L. G. McClymont and D. C. Simpson, ''Noise levels and exposure patterns to do-it-yourself power tools,'' J. Laryngol. Otol. **103**(12), 1140–1141 (1989).

[5] A. Axelsson, ''Recreational noise exposure and its effects,'' Noise Control Eng. J. **44**(3), 127–134 (1996).

[6] W. Clark, ''Noise exposure from leisure activities: a review,'' J. Acoust. Soc. Am. **90**, 175–181 (1991).

[7] D. L. Johnson and E. R. Farina, ''Description of the measurement of an individual's continuous sound exposure during a 31-day period,'' J. Acoust. Soc. Am. **62**, 1431–1435 (1977).

[8] T. Schori and E. McGatha, ''A real-world assessment of noise exposure,'' Sound Vib. **12**(9), 24–30 (1978).

[9] A. Cohen, J. Anticaglia, and H. Jones, '' 'Sociocusis:' hearing loss from non-occupational noise exposure,'' Sound Vib. **4**(11), 12–20 (1970).

[10] R. Neitzel, N. S. Seixas, J. Camp *et al.*, ''An assessment of occupational noise exposures in four construction trades,'' Am. Ind. Hyg. Assoc. J. **60**(6), 807–817 (1999).

[11] N. S. Seixas, K. Ren, R. Neitzel *et al.*, ''Noise exposure among construction electricians,'' Am. Ind. Hyg. Assoc. J. **62**(5), 615–621 (2001).

[12] C. K. Reeb-Whitaker, N. S. Seixas, L. Sheppard, and R. Neitzel, ''Accuracy of task recall for epidemiological exposure assessment to construction noise.'' Occ. Environ. Med. (in press, 2003).

[13] N. E. Klepeis, W. C. Nelson, W. R. Ott *et al.*, ''The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants,'' J. Exp. Anal Environ. Epidemiol. **11**(3), 231–252 (2001).

[14] N. Seixas, L. Sheppard, and R. Neitzel, ''Comparison of task-based estimates with full shift measurements of noise exposure,'' Ind. Hyg. Assoc. J. (in press 2003).

[15] E. Shaw, ''Noise environments outdoors and the effects of community noise exposure,'' Noise Control Eng. J. **44**(3), 109–119 (1996).

[16] A. Yassi, N. Pollock, N. Tran *et al.*, ''Risks to hearing from a rock concert,'' Can. Fam. Physician **39**, 1045–1050 (1993).

[17] T. Mori, ''Effects of record music on hearing loss among young workers in a shipyard,'' Int. Arch. Occup. Environ. Health **56**(2), 91–97 (1985).

[18] A. Axelsson, T. Jerson, and F. Lindgren, ''Noisy leisure time activities in teenage boys,'' Am. Ind. Hyg. Assoc. J. **42**, 229–233 (1981).

[19] W. Passchier-Vermeer and W. F. Passchier, ''Noise exposure and public health,'' Environ. Health Perspect. **108**(Suppl 1), 123–131 (2000).

[20] J. S. Odess, ''Acoustic trauma of sportsman hunter due to gun firing,'' Laryngoscope **82**(11), 1971–1989 (1972).

[21] D. Henderson and R. P. Hamernik, ''Impulse noise: a critical review,'' J. Acoust. Soc. Am. **80**, 569–584 (1986).

[22] D. Henderson and R. Hamernik, ''Biological bases of noise-induced hearing loss,'' Occup. Med.: State of the Art Reviews **10**(3), 513–534 (1995).

[23] M. Kerr, L. Brosseau, and C. S. Johnson, ''Noise levels of selected construction tasks,'' Am. Ind. Hyg. Assoc. J. **63**, 334–339 (2002).

[24] P. LaBenz, A. Cohen, and B. Pearson, ''A noise and hearing survey of earth-moving equipment operators,'' Am. Ind. Hyg. Assoc. J. **28**(2), 117–128 (1967).

[25] M. Legris and P. Poulin, ''Noise exposure profile among heavy equipment operators, associated laborers, and crane operators,'' Am. Ind. Hyg. Assoc. J. **59**, 774–778 (1998).

[26] S. Schneider and P. Susie, ''Final report: An investigation of health hazards on a new construction project'' (The Center to Protect Workers' Rights, Washington, DC, 1993).

[27] W. Utley and L. Miller, ''Occupational noise exposure on construction sites,'' Appl. Acoust. **18**, 293–303 (1985).

[28] A. Suter, ''Construction noise: Exposure, effects, and the potential for remediation; A review and analysis.'' AIHAJ **63**, 768–789 (2003).

Nonlinear synthesis of input signals in ultrasonic experimental setups

Frédéric Conil,[a] Dominique Gibert,[b] and Florence Nicollin[c]
Géosciences Rennes (CNRS UMR 6118), Université Rennes 1, Bât. 15 Campus de Beaulieu, 35042 Rennes cedex, France

(Received 21 April 2003; revised 12 August 2003; accepted 15 August 2003)

A method is proposed to find the digital input signals to enter in nonlinear acoustical systems (power amplifiers, transducers, etc.) in order to obtain desired arbitrary response signals. The searched input signals are found by performing a Monte Carlo search guided by a simulated annealing process applied to a hidden model with a small number of parameters. The physical system is actually used in the optimization procedure, in a real-time manner, so that no theoretical model of the system response is required. The main aspects of the algorithm are described and illustrated with several examples. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1618241]

PACS numbers: 43.58.Vb, 43.58.Ry, 43.60.Mn [EJS] Pages: 246–252

I. INTRODUCTION

The design and the control of the signals emitted and received during acoustical experiments is an important issue in many situations. Indeed, the information of interest is often contained in tiny features present in the received signals and defined with respect to reference signals recorded in calibration experiments. Such changes may, for instance, be small amplitude or phase changes, dispersion effects, time delays, etc. In all cases, the better the control of the reference signals, the better the signal-to-noise ratio and the quality of the information. A popular approach followed in many experiments consists in using impulse input signals with a large bandwidth and to performing a deconvolution of the recorded signals in order to reduce the response signals into the time responses which would be obtained for a Dirac $\delta(t)$ entered in the system. A complementary dual technique consists in entering the system with sinusoidal input signals in order to obtain the response in the frequency domain. Both approaches assume that the whole system, i.e., the acoustical apparatus plus the studied physical system, is linear so that the superposition principle is valid (e.g., Bendat and Piersol, 1971). However, there remain many instances where acoustical setups operate nonlinearly out of their nominal ranges. This is frequent in geophysics and physical oceanography where high acoustical power is necessary to insonify remote targets or deep geological layers in tomography experiments and seafloor imaging. In such cases both the piezoelectric transducers and the power amplifiers have nonlinear responses which may considerably complicate laboratory post-processing operations [Fig. 1(a)]. Classical instrumental deconvolution and mode superposition can no more be used to process these nonlinear transfer functions. Waveform distortions can also be produced by nonlinear propagation of sound in water. In some instances, these distortions may need corrections although they mainly occur for long propagation distances. Even when operating in the linear domain, deconvolution operations remain a delicate task because of their fundamental ill-posedness and of their lack of robustness with respect to noise (Jurkevics and Wiggins, 1984). Hence, in delicate high-precision experiments where instrumental responses have to be carefully controlled, deconvolution may become a limiting factor because of an insufficient numerical precision.

In recent experiments (Le Gonidec, 2001; Le Gonidec *et al.*, 2002, 2003; Gautier, 2002; Gautier and Gibert, 2003), we used an alternative approach where the nonlinear response of the whole apparatus is accounted for in a preliminary reference experiment. This is done by searching for the input signals $s(t)$ to enter into the power amplifier, triggering the acoustical source such that the corresponding response signals $r(t)$ recorded by the receiving transducers satisfy *a priori* imposed constraints (Fig. 2). These constraints may take various forms such as, for instance, a prescribed shape for $r(t)$, a maximum time duration, a controlled spectral content, an imposed regularity, etc. The principal advantage of this method is that no further processing is needed in the remaining experiments, and data with both a high signal-to-noise ratio and a good time resolution can be acquired in a real time manner. In order to account for a possibly very nonlinear distorting behavior of the system (especially the source line) the proposed algorithm uses a fully nonlinear optimization method based on simulated annealing. A key point of the method is that the whole physical system is actually included in the optimization procedure so that no theoretical model of the system response is needed. In opposition to most inverse problems whose dual direct problems are numerically solved, the forward problem considered in the present study is the physical system by itself, i.e., the whole acoustical setup composed of arbitrary waveform generators, power amplifiers, piezoelectric transducers, water tank, A/D converters, etc. By this way, the sought input signals are found by directly using the real system whatever its response, which indeed may remain unknown. This approach differs from more traditional optimization algorithms which need a model of the system response (e.g., Trucco, 2002).

[a]Electronic mail: frederic.conil@univ-rennesl.fr
[b]Author to whom correspondence should be addressed. Electronic mail: dominique.gibert@univ-rennes1.fr
[c]Electronic mail: florence.nicollin@univ-rennes1.fr

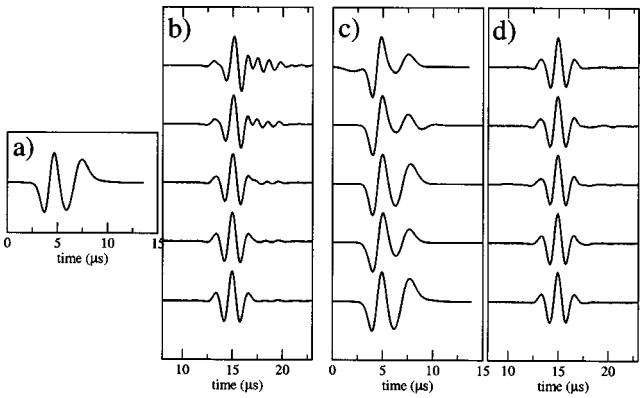


FIG. 1. Examples of nonlinear responses observed with two piezoelectric transducers immersed in a water tank (see Fig. 2). (a) Input signal sent by the arbitrary waveform generator. (b) Response signals obtained by progressively increasing (from bottom to top) the power of the source unit. The larger the source power, the more the nonlinear distortions in the late part of the received signals. (c) Input signals obtained by simulated annealing and constrained to produce response signals (d) with a prescribed shape for the same source powers as in (b). For clarity, all signals have their maximum value normalized to 1.

The high-speed electronic and computerized devices available today make it possible to plug in experimental setups into nonlinear inversion algorithms like simulated annealing. One objective of this paper is to show that nonlinear algorithms like simulated annealing, which need to solve the forward problem a huge number of times, can now be used to optimize routine experiments in a reasonable delay.

II. METROPOLIS ALGORITHM AND SIMULATED ANNEALING

A. Numerical modeling of input signals

Simulated annealing is an intensive iterative stochastic method which tests a large number of input signals $s(t)$ in order to progressively converge toward an optimal solution. The number of signals to be tested is proportional to the combinatorial complexity of the problem at hand which depends on the number of unknown parameters to optimize. For instance, a direct search of the $M=1024$ first values of an 8-bit discretized input signal results in a 256^{1024} combinatorial complexity, implying to test a too huge number of input signals to warrant a convergence toward well-optimized solutions in a reasonable delay. In the present study, we reduce the dimensionality of the optimization problem by writing the input signal as a linear combination of elementary functions $\Psi_j(t)$ parametrized with three quantities, namely, the amplitude A_j, the dilation a_j, and the time position t_j which are the unknown parameters to estimate. Practically, this reads

$$s(t|\mathbf{m})=\sum_{j=1}^{N} A_j \Psi\left(\frac{t-t_j}{a_j}\right), \tag{1}$$

where N is the total number of elementary functions used to construct the input signal. The vector $\mathbf{m}\equiv\cup_{j=1}^{N}\{A_j,a_j,t_j\}$ is the set of parameters to be optimized and is really the hidden model to be inverted. A large choice of elementary functions Ψ is possible, and in the present study we use elementary functions Ψ belonging to the class of the wavelets with constant shape (e.g., Holschneider, 1995) and defined as

$$\Psi(t)=\frac{d^n}{dt^n}\exp(-\pi t^2). \tag{2}$$

These wavelets are bandpass functions with a peak frequency depending on the value of dilation a_j: the smaller the dilation, the higher the frequency. The bandwidth is controlled by the derivation order n. Examples of wavelets corresponding to $n=1$, 2, 3, and 4 are shown in Fig. 3. In this study, wavelets with $n=1$ are used to construct the input signals. Surprisingly, even when $N\simeq 10$ in Eq. (1), a sufficiently large subspace of input signals is generated to contain the optimized input signals giving the desired response signals. The possibility to use a hidden model dramatically reduces the number of model parameters (e.g., dim $\mathbf{m}=30$ for $N=10$) and strongly speeds up the convergence toward high-probability input models.

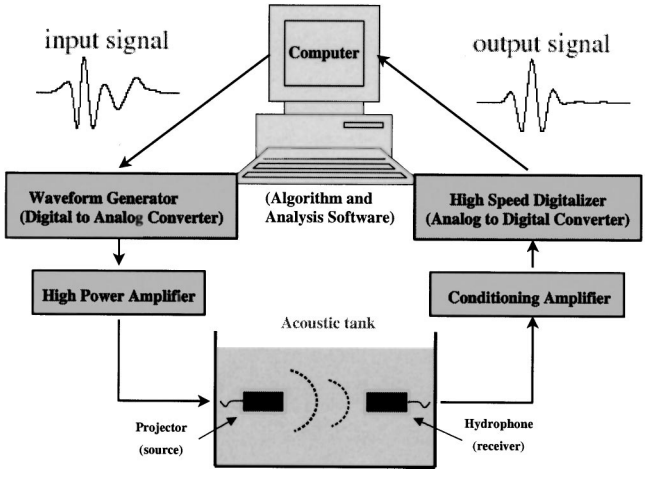


FIG. 2. Synoptic view of the experimental setup used in the present study. The source line is shown on the left part of the figure and is made of an arbitrary waveform generator, a power amplifier, and a source transducer (projector). The receiver line is shown on the right part of the figure and is made of a receiving hydrophone, a preamplifier, and an A/D converter. Both lines are connected to a computer which hosts both the simulated annealing algorithm and the driving softwares of the apparatus.

B. Probabilistic inversion

In the present paper, we consider the problem of input signal synthesis as an inverse problem formulated in a probabilistic form (Tarantola, 1987). With this approach the optimization procedure consists in searching for the models $\mathbf{m}$ with the largest posterior probability,

$$\rho(\mathbf{m})=M(\mathbf{m})S(\mathbf{m})R(\mathbf{m}), \tag{3}$$

where M is the prior probability density of $\mathbf{m}$, and S and R are probability densities which quantify the acceptability of $\mathbf{m}$ with respect to the corresponding signals $s(t|\mathbf{m})$ and $r(t|\mathbf{m})$.

When the constraints concerning the response signal $r(t|\mathbf{m})$ are given as a reference signal $r_r(t)$ to be fitted, the probability density $R(\mathbf{m})$ explicitly depends on the misfit between $r_r(t)$ and $r(t|\mathbf{m})$. The most classical definitions use the least-squares L_2 misfit,

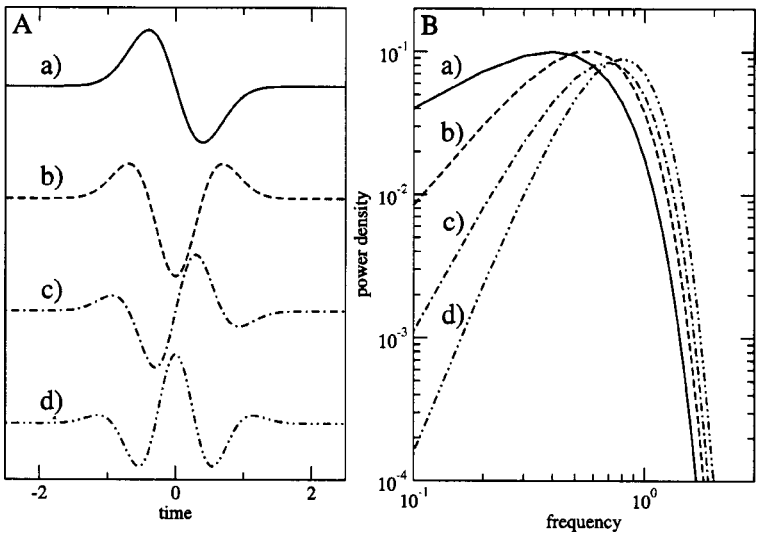


FIG. 3. A: Examples of elementary wavelets defined as the first (a), second (b), third (c), and fourth (d) derivatives of a Gaussian function. The wavelet (b) is also commonly referred to as the "*Ricker*" wavelet in seismic and as the "*Mexican hat*" wavelet in the wavelet literature. B: Power spectra of the wavelets shown on the left part.

$$R(\mathbf{m}) \equiv \eta \exp\left[-\int \frac{(r(t|\mathbf{m}) - r_r(t))^2}{2\sigma^2(t)} \mathrm{d}t \right], \tag{4}$$

or the more robust L_1 misfit,

$$R(\mathbf{m}) \equiv \eta \exp\left[-\int \frac{|r(t|\mathbf{m}) - r_r(t)|}{\sigma(t)} \mathrm{d}t \right], \tag{5}$$

where η is a normalization constant and $\sigma(t)$ is a "standard deviation" function which controls the degree of misfit tolerance. These definitions may account for variable fit constraints depending on which part of the response signal is considered. For instance, one may impose the response $r(t|\mathbf{m})$ to tightly fit the beginning of the reference signal $r_r(t)$ while a less-tight fit is acceptable in the remaining part of the signal. In other situations the constraints to be satisfied by the response signal may be defined in a more fuzzy manner so that the probability functions (4) and (5) are no more valuable. This may, for instance, be the case if the constraints on the response signal only concern its sign and its total duration T. An example of probability density accounting for these constraints may be

$$R(\mathbf{m}) \equiv \eta \exp\left[-\int_0^T [\mathrm{sign}(r(t|\mathbf{m})) - \mathrm{sign}(r_r(t))]^2 \mathrm{d}t - \int_T^\infty r^2(t|\mathbf{m}) \mathrm{d}t \right]. \tag{6}$$

In a similar manner, the probability function $S(\mathbf{m})$ is used to measure the acceptability of $\mathbf{m}$ with respect to its input signal $s(t|\mathbf{m})$. Contrary to the case of the response signal on which tight constraints can be applied, the precise form of the input signal is generally not important and the constraints are less hard. These may, for instance, be bounds on the amplitude of the signal or imposed limits on the total dissipated power in order to protect the apparatus against overload and failure.

The prior probability $M(\mathbf{m})$ may be used to control the complexity of the model and, in particular, to reduce the nonuniqueness of the solution and eliminate the components Ψ_j of $s(t|\mathbf{m})$ which belong to the null-space of the physical system, i.e., such that they produce a null response. As will be seen in a next section, reducing the nonuniqueness may be of some interest to construct a family of input signals further used to train neural networks. Indeed, the presence of unnecessary components in the optimized model results in a randomization of the training ensemble which may considerably alter the convergence of the neural net parameters during the training phase. In the present study, when used, the prior probability is simply defined as the inverse of the number of non-null components of the model $\mathbf{m}$,

$$M(\mathbf{m}) = \frac{1}{\dim\{A_j \neq 0\}}. \tag{7}$$

Our probabilistic formulation of signal synthesis allows us to freely define the imposed constraints to be satisfied by both the input and the response signals. This freedom results in a highly nonlinear influence of the model parameters which, together with the fact that the response of the physical apparatus is unknown and also nonlinear, eliminates the possibility to use classical optimization methods like steepest descent. Furthermore, these methods generally involve the Hessian matrice or the Fréchet's matrice of the partial derivatives $\partial r(t|\mathbf{m})/\partial m_k$ which are not easy (but not impossible) to evaluate for physical systems.

C. The Metropolis algorithm

The Metropolis algorithm (Metropolis *et al.*, 1953) is a Markov Monte Carlo chain which generates a sequence of models $\{\mathbf{m}_l\}$ distributed according to a prescribed probability density function ρ. The sequence is constructed as follows. Let $\mathbf{m}_l$ be the last model in the sequence, and let $\mathbf{m}_?$ be a candidate model to eventually become the next model $\mathbf{m}_{l+1}$. Then, the candidate model is retained and enters into the Metropolis sequence according to the following stochastic rules:

$$\mathbf{m}_{l+1} = \mathbf{m}_? \quad \text{with probability} \max\left[1, \frac{\rho(\mathbf{m}_?)}{\rho(\mathbf{m}_l)}\right], \tag{8}$$

$$\mathbf{m}_{l+1} = \mathbf{m}_l \quad \text{instead.} \tag{9}$$

These equations indicate that the candidate model $\mathbf{m}_?$ is eventually accepted with probability $\rho(\mathbf{m}_?)/\rho(\mathbf{m}_l)$ if $\rho(\mathbf{m}_?)$

$<\rho(\mathbf{m}_l)$ and is systematically accepted if $\rho(\mathbf{m}_?)\geqslant\rho(\mathbf{m}_l)$. The Metropolis sequence is iteratively constructed by randomly generating candidate models and by using the selection formulas (8) and (9). It can be shown that the models forming the so-constructed sequence are sampled according to the probability density ρ (Bhanot, 1988).

D. Simulated annealing

Simulated annealing (Kirkpatrick *et al.*, 1983) is a global search algorithm aimed at converging toward high-probability models for nonconvex, and even fractallike, posterior probability functions. The basic idea of simulated annealing is to iterate the Metropolis algorithm while controlling the topology of the posterior probability density function in order to progressively transform the sequence of models from a uniformly sampled one to a sequence of models confined in the vicinity of the global maximum of the posterior probability function. Since the precise topological characteristics of the probability density remain unknown, the topological deformation must rely on very general principles. For the simulated annealing algorithm, this is done through the use of a control parameter T, called the temperature for historical reasons, which is varied from infinity to 1. Practically, the topological transformation of the posterior probability reads

$$\rho_T(\mathbf{m})=\frac{1}{G}\rho^{1/T}(\mathbf{m}), \tag{10}$$

where ρ_T is the transformed posterior probability density for temperature T, and G is a normalization constant (in fact the partition function). Asymptotically ρ_∞ tends toward the uniform probability density, and for $T\downarrow 0$ we have $\rho_0\rightarrow\delta(\mathbf{m}-\mathbf{m_{MP}})$ where $\mathbf{m_{MP}}$ is the model with the maximum posterior probability.

Simulated annealing consists in iterating the Metropolis algorithm while decreasing the temperature T from a high value to 1. Provided the temperature is sufficiently slowly decreased, the sequence of models $\{\mathbf{m}_j\}$ progressively concentrates in the regions with high posterior probability. For this guidance to be efficient, some correlation is necessary among the nearby models of the sequence in order to perform a random walk rather than a random sampling in the model space. This is particularly true at low temperature where a pure Monte Carlo sampling would be very inefficient because of a too high rate of rejection of the tested models in the Metropolis loop. In the present study, the correlation among the models is done by two ways. First, as usually done in simulated annealing, the tested models $\mathbf{m}_?$ are generated by stochastically perturbing the last accepted model $\mathbf{m}_j$, i.e., by only modifying a subset of the model parameters. Second, the range of allowed changes for the model parameters is narrowed as the temperature decreases. This approach proved particularly efficient when dealing with continuous variables, as in the present study (Vanderbilt and Louie, 1984). In the present study, we implemented this method by performing a statistical analysis of the model sequence generated at each temperature during the annealing process. From this analysis the ranges of variation of the model parameters are progressively narrowed and centered on the region of high posterior probability $\rho(\mathbf{m})$ where the models cluster. This procedure considerably accelerates the convergence. Since the Metropolis uses the ratio of probabilities [see Eqs. (8) and (9)], the normalization constant G in Eq. (10) may be omitted and the probability distribution of the models in the Metropolis sequence may be *a posteriori* normalized.

III. EXPERIMENTAL NONLINEAR SYNTHESIS OF INPUT SIGNALS

A. Experimental setup

The experimental setup is shown in Fig. 2 and is composed of electronic devices, piezoelectric transducers, and a water tank of 5 m^3. All devices are remotely controlled by a master software written in LabView® language. First, the last model $\mathbf{m}_l$ of the Metropolis sequence is perturbed to give the candidate model $\mathbf{m}_?$ used to compute a discretized version of the input signal $s(t|\mathbf{m}_?)$ which is sent into the arbitrary waveform generator. Next, the analog physical electrical signal corresponding to s is created by the generator and entered in the power amplifier connected to the source transducer which produces an acoustical wave. Next, the acoustical wave reaches the piezoelectric receiver which is connected to a preamplifier and to a digital oscilloscope which sends a discretized version of $r(t,\mathbf{m}_?)$ to the computer. Finally, the response signal is used to compute $R(\mathbf{m}_?)$ and the posterior probability $\rho(\mathbf{m}_?)$ used in the Metropolis decision rules (8) and (9). After either $\mathbf{m}_?$ or $\mathbf{m}_l$ became the new last model $\mathbf{m}_{l+1}$ in the Metropolis sequence, a new candidate model is again created and the whole procedure is repeated. A real-time control of the simulated annealing process is done by plotting $r(t,\mathbf{m}_{l+1})$, $r_r(t)$, and $\rho(\mathbf{m}_{1,\ldots,l+1})$ in a LabView® window. This display of the intermediate results of the optimization process is useful to observe the efficiency of the cooling schedule in the simulated annealing loop.

B. Synthesis of input signals for an imposed response

In this first example, several input signals are inverted to produce a unique given response signal at a fixed source power. Figures 4(a) and (b) show the results of five inversions obtained with the constraint on the response signal given as an imposed reference signal $r_r(t)$ chosen to be the wavelet shown in Fig. 3(d). Both $M(\mathbf{m})=1$ and $S(\mathbf{m})=1$ in this example and the probability density $R(\mathbf{m})$ has the form given by Eq. (4) with a constant σ. As can be seen, the obtained input signals have conspicuous differences [Fig. 4(a)] while their corresponding responses remain identical [Fig. 4(b)]. These results illustrate the fact that the physical system, i.e., the forward problem in the inverse problem terminology, possesses a nonempty null-space containing input signals whose response is zero. These null-space elements have a frequency content so out-of-bandwidth that no response signal is produced by the physical system. Another form of nonuniqueness, but nonvisible in this example, is due to the fact that a single wavelet in the synthesis formula (1) may be replaced by two or more wavelets having the same dilation a_j and time position t_j. This first example

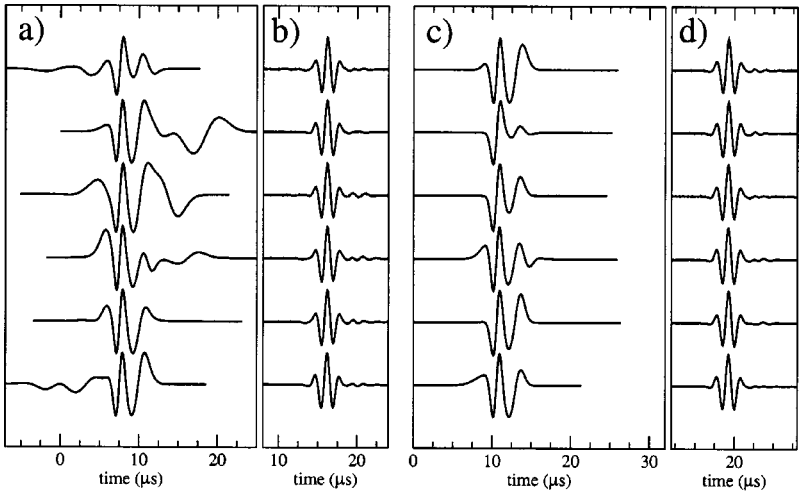


FIG. 4. Results of inversions performed for a unique given reference signal and a fixed source power. (a) Input signals inverted with both $M(\mathbf{m})=1$ and $S(\mathbf{m})=1$. (b) Corresponding response signals. (c) Input signals inverted with $M(\mathbf{m})$ given by Eq. (7). (d) Corresponding response signals.

shows that a given arbitrary response signal may be obtained with a good quality. The fact that this response may be produced by rather different input signals is generally not a problem and illustrates the ill-posedness of deconvolution. However, there may be particular instances where it is desirable to reduce nonuniqueness when inverting a family of input signals. Figure 4(c) shows the inverted signals equivalent to those of Fig. 4(a) but obtained by using the prior probability $M(\mathbf{m})$ given by Eq. (7). As can be seen, this additional constraint strongly reduces the discrepancies among the inverted input signals without altering the quality of the corresponding response signals [Fig. 4(d)]. Slight discrepancies between the output signals may be observed particularly in their coda parts where tiny wiggles appear in some outputs. These oscillations remain in the obtained solutions because the input model $\mathbf{m}$ has a limited number of wavelets Ψ_j which concentrate in the early part of the input signal $s(t|\mathbf{m})$ in order to correctly reproduce the main part of the desired output signal. If unwanted, these artifacts could be suppressed by using a more complicated input model with a sufficiently large number of wavelets Ψ_j so that a longer and more adapted input signal could be constructed by simulated annealing.

Figure 5 shows the annealing process observed during an inversion. The posterior probability $\rho(\mathbf{m})$ of the models accepted during the Metropolis loops may be very low when the temperature parameter is high at the beginning of the iterative procedure. This stage of the annealing is critical for it allows a global search among the model space and important changes of the parameter values. This is possible because at this stage of the process the random walk in the model space may go through very low probability models in order to eventually escape from local maxima of the posterior probability. It is important that the global search can quickly explore large parts of the model space in order to localize the region with the global maximum of probability. This region is identified by performing a statistical analysis of the models accepted during each successive Metropolis loop in order to see where the accepted models progressively cluster. Once the region with the global maximum is identified, the posterior probability of the accepted models drops suddenly when the temperature decreases [around iteration number 20 in Fig. 5(a)], and we observe that the remaining iterations of the annealing procedure are restricted to a subregion of the model space. It is, however, necessary to continue the annealing procedure with a sufficiently slow cooling in order to gently guide the model sequence toward the global maximum. We observe that a too-rapid decrease of the temperature traps the models into a local maximum.

C. Synthesis of input signals for different source powers

The examples presented in this section concern the synthesis of the input signals producing a given response wavelet for the same power levels as in Figs. 1(a) and (b). As in the preceeding section, the constraint on the response signal is given as a reference signal and the nonuniqueness is reduced by using the prior probability $M(\mathbf{m})$ given by Eq. (7). The inverted input signals are shown in Fig. 1(c) and display subtle differences to account for the nonlinearities of the physical system. The corresponding response signals are shown in Fig. 1(d) and look very identical as expected.

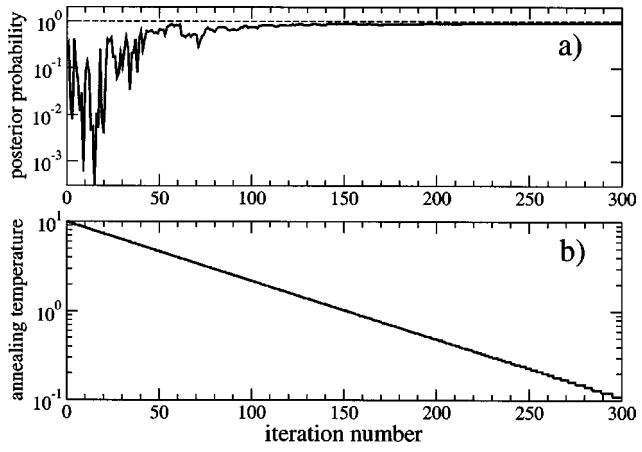


FIG. 5. (a) Posterior probability $\rho(\mathbf{m})$ of the models accepted during an annealing sequence with a temperature schedule given by $T_{k+1}=0.985T_k$ in part (b) of the figure. Each iteration counts 20 model tests at a fixed temperature, so the 300 iterations shown here correspond to 6000 tested (and eventually accepted) models. In practice, this takes about 10 min to be performed.

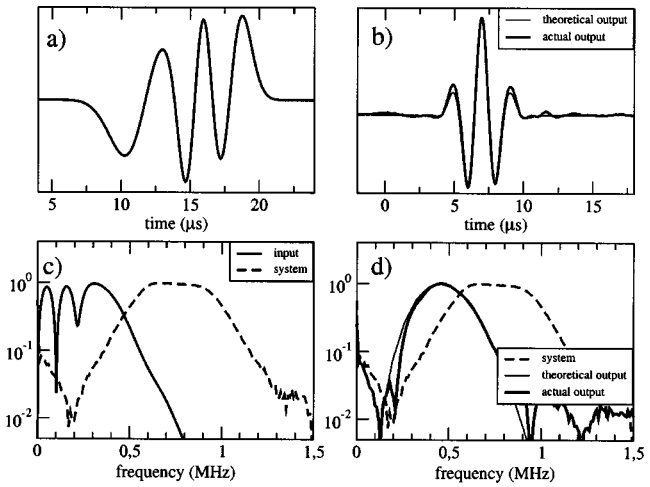


FIG. 6. Synthesis of input signals out of the main frequency band of the system. (a) Optimized input signal. (b) Actual response signal and reference response signal. (c) Spectral amplitude of the optimized input signal shown in (a). (d) Spectral amplitude of the response signals (both actual and reference) and of the impulse response of the physical system.

D. Synthesis out of the main frequency band

An interesting feature of the proposed algorithm is that it allows us to find input signals whose response signals have a frequency content partly out of the nominal bandwidth of the physical system. This possibility is particularly useful to construct a family of response signals covering a wide frequency bandwidth. Even when the physical system operates in a linear regime, classical deconvolution methods using an impulsive input signal do not produce satisfactory results concerning the out-of-bandwidth responses. This is because most of the energy of the input impulse is released at the central bandwidth of the system, rejecting the useful bandwidth at the noise level. A great interest of our method is that the input signals are constructed in such a manner that energy is sent only at the relevant frequencies so that no saturation occurs due to strong energy release in the dominant frequency band. Figure 6 shows the results of such an inversion. One can observe that the algorithm is able to find an input signal whose corresponding response has a spectral content out of the central bandwidth of the physical system. This can be obtained by producing an input signal with almost no energy located in the central band and such that a high signal-to-noise ratio is obtained.

E. Synthesis of a wavelet family

The proposed method allows us to generate families of signals optimally located in the time-frequency domain, as shown in Fig. 7, where the response signals are dilated versions of the wavelet (d) in Fig. 3. In this example, four pairs of transducers with central frequencies of 250, 500, 750, and 1000 kHz were used, and the frequency range covered by the whole set of signals goes from 150 kHz to 1.2 MHz. Such a family of signals is useful to study the frequency response of a part of a complex system. Indeed, in this kind of situation, it is necessary to extract the response of the studied system element from a complex and intricate global response. This extraction is generally impossible when the frequency response is studied through the classical Fourier approach by using sinusoidal input signals with a long duration. Conversely, using a wavelet family as shown in Fig. 7 allows us to obtain an ensemble of responses of the system with good time-frequency properties. The time localization of the signals enables a temporal separation of the response of subsystems of interest while controlling the time-frequency characteristics of the signals.

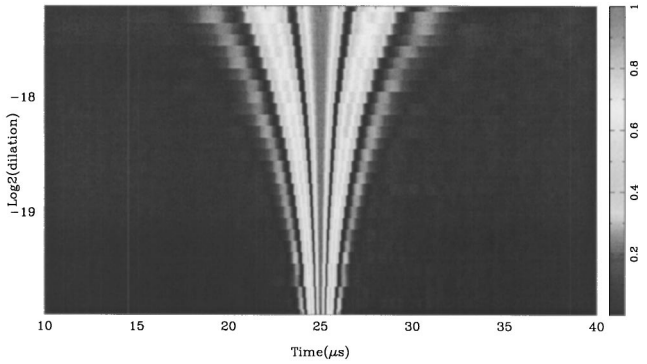


FIG. 7. Output wavelet family corresponding to a collection of inverted input signals constrained to produce dilated versions of the wavelet shown in Fig. 3(d). Each line of the color array represents a wavelet whose absolute value is coded according to the color scale shown on the right. The dilations are such that $a=f_c^{-1}$ where f_c is the central frequency of the response signal given in Hz.

F. Application: Reflectivity study of granular media

We now give an example of application of the wavelet family shown in Fig. 7. The objective of the experiment is to study the frequency dependence of the reflectivity of the surface of a layer of disordered monodisperse glass beads (see Le Gonidec *et al.*, 2002, 2003 for details). The experiment consists in recording the spatially averaged wavefield reflected by the surface in response to the incident wavelets of Fig. 7. Figure 8 shows the spatially averaged reflected signals corresponding to the input signals of the wavelet family. The time origin of each averaged signal has been arbitrarily shifted in order to align the second extrema of all signals. At large dilations (i.e., large wavelengths) the reflected signals are arranged as a conical structure which converges toward the average time-location of the surface of the heterogeneous

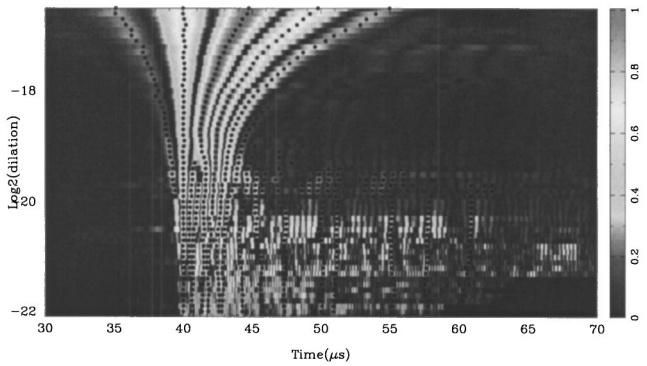


FIG. 8. Modulus of the average coherent signal reflected by a layer of glass beads in response to the incident wavelets shown in Fig. 7. The dilations are such that $a=f_c^{-1}$ where f_c is the central frequency of the response signal given in Hz.

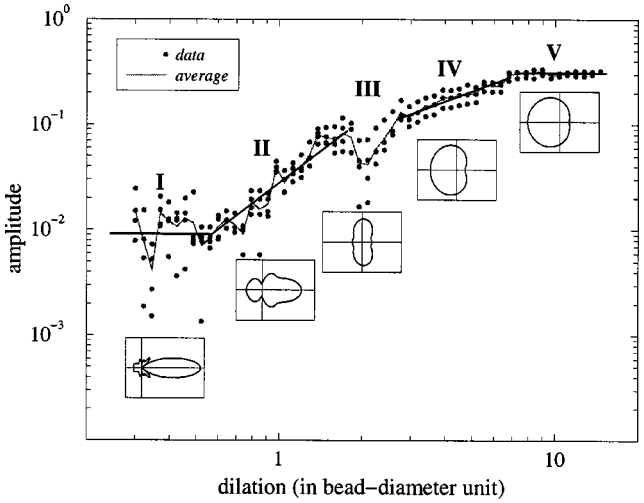


FIG. 9. Reflectivity of the layer of glass bead inferred from the amplitude of the coherent reflected signals shown in Fig. 8. See Le Gonidec *et al.* (2002, 2003) for details.

layer. At smaller dilations (i.e., shorter wavelengths) the main cone of the wavelet response is followed by a coda, and the time localization of the reflector is blurred. The frequency dependence of the reflectivity may be accurately studied by plotting the amplitude of the reflected wavelets against the wavelength normalized by the bead diameter (Fig. 9). At wavelengths larger than seven times the bead diameter (waveband V in Fig. 9) the reflectivity is constant and the heterogeneous layer behaves like a homogeneous elastic effective medium, while at smaller wavelengths (waveband IV) some attenuation occurs due to increasing scattering. The waveband III marks a transition and corresponds to a sharp decrease of the reflectivity. Physically, this narrow waveband corresponds to dominant transverse Mie scattering while backscattering dominates in bands IV and V. The reflectivity decrease in band III may then be explained by the fact that a significant fraction of the incident wave is laterally scattered out of the coherent reflected wave. Band II displays a sharper decrease of the reflectivity when the wavelength shortens. This corresponds to a dominant forward Mie scattering where more and more energy is scattered forward in the heterogeneous layer, contributing to the conspicuous coda visible in Fig. 8. In band I corresponding to the shortest wavelengths, equal to or shorter than the bead radius, the coherent reflected signal is caused by the water–glass interface, and the reflectivity is more the one of the rough water–glass interface than the one of the heterogeneous as a whole (Le Gonidec *et al.*, 2002, 2003).

We emphasize that these experimental results are directly obtained without postprocessing deconvolution. Once the wavelet family is obtained, the corresponding input signals are sent to the source chain and the reflected signals are averaged and recorded as shown in Fig. 8. The methods allows for a tight control of the shape of the desired response signals so that the reflectivity may be accurately determined.

IV. CONCLUSION

The nonlinear method presented in this paper performs a Monte Carlo search of the input signals able to produce a desired signal at the response of a nonlinear acoustical chain. This is done by simulated annealing which guides a random walk through the model space. This may be done by embedding the physical system in the operating software by using fast A/D converters. The main advantages of the proposed method are (1) the possibility to generate a unique given signal for various source power by accounting for the nonlinear distortions occurring in the physical system (Fig. 1); (2) the possibility to generate response signals with a frequency content located out of the nominal bandwidth of the system (Fig. 6); and (3) the possibility to perform an optimal time-frequency analysis and to isolate individual responses from composite responses involving different physical phenomena (Fig. 8).

ACKNOWLEDGMENTS

Yves Le Gonidec kindly gave us the data of the wavelet response example. Arnaud Derode gave us much valuable advice and generously shared his experimental and theoretical knowledge. This work is financially supported by the CNRS and ANDRA through the GdR FORPRO (Research Action No. 99.II) and corresponds to the GdR FORPRO Contribution No. 2003/02 A.

Bhanot, G. (**1988**). ''The Metropolis algorithm,'' Rep. Prog. Phys. **51**, 429–457.

Bendat, J., and Piersol, A. G. (**1971**). *Random Data: Analysis and Measurement Procedure* (Wiley Interscience, New York).

Gautier, S. (**2002**). ''Imagerie Acoustique de Discontinuités Rugueuses: Expérimentation et Modélisation Numérique'' (''Acoustical Imaging of Rough Discontinuities: Experiments and Numerical Modeling''), Ph.D. thesis, University Rennes 1.

Gautier, S., and Gibert, D. (**2003**). ''Scattering from a fractal surface: Acoustical experiments and comparison with near-nadir models,'' Icarus (in press).

Holschneider, M. (**1995**). *Wavelets: An Analysis Tool* (Clarendon, Oxford, England).

Jurkevics, A., and Wiggins, R. (**1984**). ''A critique of seismic deconvolution methods,'' Geophysics **49**, 2109–2116.

Kirkpatrick, S., Gellatt, C. D., and Vecchi, M. P. (**1983**). ''Optimization by simulated annealing,'' Science **220**, 671–680.

Le Gonidec, Y. (**2001**). ''Caractérisation d'Interfaces et de Milieux Complexes par Analyse Sismique Multi-Échelle: Expérimentation et Modélisation Numérique'' (''Interfaces and Complex Media Characterization by Multi-scale Analysis: Experiments and Numerical Modeling''), Ph.D. thesis, University Rennes 1.

Le Gonidec, Y., Conil, F., and Gibert, D. (**2003**). ''The wavelet response as a multiscale NDT method,'' Ultrasonics, **41**, 487–497.

Le Gonidec, Y., Gibert, D., and Proust, J.-N. (**2002**). ''Multiscale analysis of waves reflected by complex interfaces: Basic principles and experiments,'' J. Geophys. Res. **107**(B9), 2184.

Metropolis, N., Rosenbluth, A., Rosenbluth, N., Teller, A., and Teller, E. (**1953**). ''Equation of state calculations by fast computing machines,'' J. Chem. Phys. **21**, 1087–1092.

Tarantola, A. (**1987**). *Inverse Problem Theory* (Elsevier, New York).

Trucco, A. (**2002**). ''Weighting and thinning wide-band arrays by simulated annealing,'' Ultrasonics **40**, 485–489.

Vanderbilt, D., and Louie, S. G. (**1984**). ''A Monte Carlo simulated annealing approach to optimization over continuous variables,'' J. Comput. Phys. **56**, 259–271.

Uncertainties associated with parameter estimation in atmospheric infrasound arrays[a)]

Curt A. L. Szuberla[b)] and John V. Olson
Geophysical Institute, University of Alaska Fairbanks, Fairbanks, Alaska 99775-7320

(Received 25 April 2003; revised 3 October 2003; accepted 30 October 2003)

This study describes a method for determining the statistical confidence in estimates of direction-of-arrival and trace velocity stemming from signals present in atmospheric infrasound data. It is assumed that the signal source is far enough removed from the infrasound sensor array that a plane-wave approximation holds, and that multipath and multiple source effects are not present. Propagation path and medium inhomogeneities are assumed not to be known at the time of signal detection, but the ensemble of time delays of signal arrivals between array sensor pairs is estimable and corrupted by uncorrelated Gaussian noise. The method results in a set of practical uncertainties that lend themselves to a geometric interpretation. Although quite general, this method is intended for use by analysts interpreting data from atmospheric acoustic arrays, or those interested in designing and deploying them. The method is applied to infrasound arrays typical of those deployed as a part of the International Monitoring System of the Comprehensive Nuclear-Test-Ban Treaty Organization. [DOI: 10.1121/1.1635407]

PACS numbers: 43.60.Fg, 43.60.Cg, 43.28.Dm [EJS] Pages: 253–258

I. INTRODUCTION

A significant increase in the use of atmospheric acoustic arrays has been brought about by the deployment of a global network of infrasound sensor arrays as a part of the International Monitoring System (IMS) of the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO).[1] These IMS arrays are to be capable of observing a nuclear event of nominal yield ~1 kT in the frequency band $f \in [0.1,5]$ Hz. Critical to the detection of such an event are precise estimates of direction-of-arrival (θ) and trace velocity (v). Since precise determination of θ and v from infrasound arrays is complicated by the ubiquitous presence of wind noise and atmospheric anisotropies, recent theoretical[2,3] and experimental[4] effort has been focused on mitigation of these effects. Analysis of propagation path properties ultimately requires that the source location is known, in which case estimation of θ becomes moot. Such analyses are, however, quite useful in that they assess the performance of arrays and signal-processing techniques against known standards.

This paper develops a method for determining the statistical confidence in estimates of θ and v for those situations in which the signal source location is unknown, as are the atmospheric properties between the source and the array. This method would be particularly useful in the event of a clandestine atmospheric nuclear test. Considering the instruments available to the IMS (seismic, infrasound, hydroacoustic, and radio-nuclide), such an event would likely best be observed by an infrasound array. The method developed is ideally suited to realistic first-detection scenarios, in which the particulars of a suspected event beyond the infrasound array data are unknown, but may be determined later via analysis of other, associated data.

The method is developed in Sec. II and Sec. III under the assumption that a signal from a single source impinges upon the array as a plane wave. An ensemble of time delays between array sensor pairs is assumed to be estimable, typically by finding the lags corresponding to the maxima of a series of normalized cross-correlation function estimates, and that such time delays are corrupted by uncorrelated Gaussian noise. This delay information is then incorporated into a model equation, which is solved in the least-squares sense for an estimate of the slowness (reciprocal velocity) of the signal as it crosses the array. From this solution, the statistics of θ and v are determined. In Sec. IV the method is applied to IMS-type arrays, allowing for an analysis of the effectiveness of the array geometry in general, or for a particular signal.

While more sophisticated techniques exist[5,6] for improving the accuracy and/or precision of source localization, the use of time delays and least-squares analysis remains robust, computationally efficient, and typically represents the first line of analysis by many groups.[7–9] This technique is particularly effective in the analysis of impulsive events, such as those of interest to the CTBTO. Moreover, such methods have been used extensively in the past to study continuous, narrow-band signals such as microbaroms.[10,11]

Ultimately, development of this method is aimed not at improving the precision of source localization; rather, it is constructed to produce a set of practical uncertainties in θ and v that is useful to the analyst of infrasound array data. These uncertainties lend themselves to a geometric interpretation and serve as valuable input parameters to source location methods such as triangulation with other infrasound arrays and atmospheric propagation models. Too, they govern the uncertainty in such postanalyses.

[a)]Portions of this work were presented in "An investigation of the uncertainties associated with estimates of velocity and direction-of-arrival of infrasound signals," Proceedings of Infrasound Technology Workshop, De Bilt, The Netherlands, 28–31 October 2002.

[b)]Electronic mail: cas@gi.alaska.edu

II. SIGNAL MODEL

Detection of a nuclear excursion resulting in a nonplanar arrival at an infrasound array would not present a significant signal-processing challenge; therefore, the plane-wave assumption is particularly warranted in the case of the IMS. In this paper, the compact *bra-ket* notation of Dirac[12] for linear algebra is followed. The term trace velocity is used here for v as the IMS arrays are all roughly planar, so only the component of velocity parallel to the plane of the array may be estimated. In this section, however, the method is developed for an array of arbitrary dimension. Throughout the discussion, the $\mathbb{I}$ symbol is used to denote the identity matrix of the appropriate dimension.

Consider a d-dimensional array of n sensors. There exist $N=n(n-1)/2$ unique interstation separations which may be encoded in an $[N,d]$ matrix $\mathbb{X}$ of Cartesian coordinates. It is important to consider all of these separations in order to considerably reduce the variance in the estimates of the direction-of-arrival and velocity of the plane wave across the array. The arrival of such a plane-wave signal, whether broadband or monochromatic, gives rise to an N-dimensional vector of interstation time delays $|\tau\rangle$, which can be used to solve the model equation

$$|\tau\rangle=\mathbb{X}|s\rangle+|\epsilon_\tau\rangle, \tag{1}$$

for $|s\rangle$, the d-dimensional estimate of the slowness vector. Equation (1) is solved subject to the constraint that the sum-squared error in $|\tau\rangle$, $\langle\epsilon_\tau|\epsilon_\tau\rangle$, be minimized. The actual signal is represented by the theoretical slowness vector $|s_o\rangle$ and its associated time delay by $|\tau_o\rangle$.

Under the assumption that the errors in measuring the time delays $|\epsilon_\tau\rangle$ are normally distributed about $|\tau_o\rangle$, the covariance of $|\tau\rangle$ is uncorrelated

$$\mathbb{C}_\tau=\sigma_\tau^2\mathbb{I}. \tag{2}$$

Our assumption is only strictly valid in the bandlimited case, as turbulence has been shown to significantly affect the character of noise distributions.[3] That said, Eq. (2) appears to hold in studies we have conducted using mine explosions and the IMS array located at the University of Alaska Fairbanks. We find that simple bandpass filtration is often sufficient to resolve even weak impulsive signals in noise to within the uncertainties developed in the next section. In this case, the least-square estimator in Eq. (1) is the maximum likelihood estimator and the Cramér–Rao bound is met for the variance of $|s\rangle$.[13] Beyond simply determining the Cramér–Rao bound, which as a lower bound is intrinsically optimistic in situations where the model assumptions are not met, this paper explores the d-dimensional nature of the uncertainty in $|s\rangle$ in terms of an arbitrary confidence limit.

III. STATISTICAL MODEL

A. Distribution of time delays

Solving Eq. (1) for $|s\rangle$ in the least-squares sense gives solutions as

$$|s\rangle=\mathbb{C}^{-1}\mathbb{X}^\dagger|\tau\rangle, \tag{3}$$

where

$$\mathbb{C}=\mathbb{X}^\dagger\mathbb{X} \tag{4}$$

is the covariance matrix of interstation separations. An estimate of the variance in $|\tau\rangle$ is given by[14]

$$\sigma_\tau^2=\frac{\langle\tau|(\mathbb{I}-\mathbb{R})|\tau\rangle}{N-r}, \tag{5}$$

where

$$\mathbb{R}=\mathbb{X}\mathbb{C}^{-1}\mathbb{X}^\dagger, \tag{6}$$

and r is the rank of $\mathbb{R}$ (always d for d-dimensional arrays, the dimension of $|s\rangle$). Note here that the estimate of the variance in $|\tau\rangle$ is governed by the geometry of the array, embedded in the matrix $\mathbb{R}$. Equation (5) can be thought of as a measure of how well the observed $|\tau\rangle$ corresponds to a physically realizable set of delays across the array.

Since the delay vector $|\tau\rangle$ is assumed to be normally distributed about $|\tau_o\rangle$ in the space of interstation separations given in $\mathbb{X}$, its distribution function is given by[15]

$$f(|\tau\rangle)=\frac{1}{(2\pi)^{N/2}|\mathbb{C}_\tau|^{1/2}}e^{-(1/2)(\langle\tau|-\langle\tau_o|)\mathbb{C}_\tau^{-1}(|\tau\rangle-|\tau_o\rangle)} \tag{7}$$

$$=\frac{1}{(2\pi\sigma_\tau^2)^{N/2}}e^{-(1/2\sigma_\tau^2)(\langle\tau|-\langle\tau_o|)\mathbb{I}(|\tau\rangle-|\tau_o\rangle)}. \tag{8}$$

B. Distribution of slowness estimates

Defining the residuals of the slowness vector as

$$|\tilde{s}\rangle=|s\rangle-|s_o\rangle, \tag{9}$$

and substituting Eq. (1) into Eq. (8) gives a transformation from the N-dimensional Gaussian distribution function of $|\tau\rangle$ to the d-dimensional distribution function of $|s\rangle$

$$f(|s\rangle)=\kappa e^{-(1/2\sigma_\tau^2)\langle\tilde{s}|\mathbb{C}|\tilde{s}\rangle}, \tag{10}$$

which is also Gaussian. Note that the exact value of the normalization constant κ in Eq. (10) is unimportant. Surfaces of constant probability on the distribution function are of interest here. In the case of Gaussian distributions the surfaces are ellipsoids, and this method ascertains the uncertainties in $|s\rangle$ via these concentration ellipsoids.[15]

The covariance matrix of interstation separations may be diagonalized via the eigenvalue equation

$$\mathbb{D}=\mathbb{E}^\dagger\mathbb{C}\mathbb{E}, \tag{11}$$

where $\mathbb{D}$ is the $[d,d]$ diagonal matrix of eigenvalues and $\mathbb{E}$ is the $[d,d]$ matrix of eigenvectors of $\mathbb{C}$. Substituting Eq. (11) into Eq. (10) gives

$$f(|s\rangle=\kappa e^{-(1/2\sigma_\tau^2)\langle\tilde{s}|\mathbb{E}\mathbb{D}\mathbb{E}^\dagger|\tilde{s}\rangle} \tag{12}$$

$$=\kappa e^{-(1/2\sigma_\tau^2)\langle\hat{s}|\mathbb{D}|\hat{s}\rangle}, \tag{13}$$

where

$$|\hat{s}\rangle=\mathbb{E}^\dagger|\tilde{s}\rangle \tag{14}$$

is seen to be the result of a rotation of coordinates of the residuals of the slowness vector into the principal axes of the covariance matrix of the interstation separations.

Geometrically then, setting the exponent in Eq. (13) to various constants gives constant probability surfaces of the distribution function of the slowness vector estimate $|s\rangle$. Choosing the constant to be $\Delta\chi^2_{\nu|p}$ (χ^2 with ν degrees of freedom at confidence level p), it is evident that the exponent in Eq. (13) defines the equation of a d-dimensional ellipsoid

$$\Delta\chi^2_{\nu|p}=\sum_{i=1}^{d}\frac{\hat{s}_i^2}{\sigma^2_{\hat{s}_i}}, \tag{15}$$

which is centered at $|s_o\rangle$ and whose principal axes are aligned with the eigenvectors of $\mathbb{C}$, with lengths given by

$$\sigma_{\hat{s}_i}=\sqrt{\frac{\sigma_\tau^2}{D_{ii}}}. \tag{16}$$

For a given ensemble of estimates $|s\rangle$ of $|s_o\rangle$ from Eq. (3), a fraction p of them will be bounded by the ellipsoid.[16]

C. Geometric interpretation of uncertainty

Once the form of the concentration ellipsoid in $|s\rangle$ is known, the uncertainty may be determined to any desired confidence. There exists, however, an alternative means of describing the uncertainty in the special case where $d=2$ (planar arrays). This method is of more practical use to the analyst of IMS infrasound data than a statistical description of uncertainties along the eigenvectors of the covariance matrix of interstation separations. In this case, Eq. (15) takes on the familiar form of an ellipse centered at $|s_o\rangle$

$$\Delta\chi^2_{\nu|p}=\frac{\hat{s}_x^2}{\sigma^2_{\hat{s}_x}}+\frac{\hat{s}_y^2}{\sigma^2_{\hat{s}_y}}. \tag{17}$$

Although many of the IMS arrays are nonplanar, we have found that if an array's aperture is at least an order of magnitude greater than its vertical relief, then the uncertainties associated with nonplanar geometry are negligible. For array placements in more exaggerated terrain, we recommend numerically computing bounds similar to the ones described below using the full form of Eq. (15).

Ignoring ellipses that encompass the origin, indicative of simultaneous arrival across the entire array, the ellipse of Eq. (17) subtends a finite angle from the origin; too, it occupies a finite radial expanse. A measure of the uncertainty in $|s\rangle$ may then be defined as the angle subtended by the ellipse ($\delta\theta$) from the origin, and the difference between extremal distances from the origin to the ellipse (δs). This uncertainty measure is conservative, as it defines the area along an arc segment between two concentric circles, a locus of points of which the concentration ellipse is a subset. The angle subtended by the ellipse may be found either analytically or numerically by solving a quadratic equation for the angle defined by the two lines tangent to the ellipse that pass through the origin. Similarly, the extremal distances to the ellipse may be found by solving a quartic equation for the lengths of the line segments which are perpendicular to tangent lines along the ellipse at their intersection and which include the origin. Since there are four possible solutions, the longest and shortest real segments are then selected.

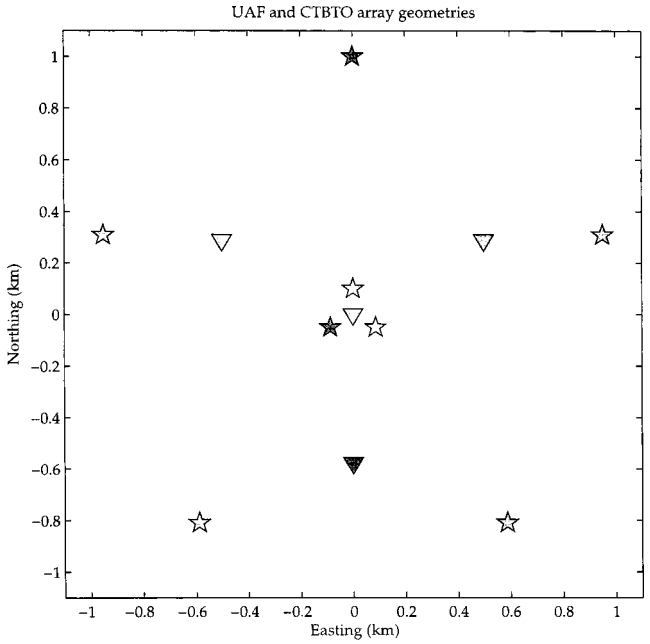


FIG. 1. IMS infrasound array geometries: UAF (pentagrams) and CTBTO (triangles). Darkened symbols represent sensors which will be disabled to study the impact of sensor failure on array performance.

Trace velocities are of more use to infrasound analysts than slownesses, so the slowness ellipse may be conformally mapped to velocity space via the following:

$$v_x=\frac{s_x}{s_x^2+s_y^2}, \tag{18a}$$

$$v_y=\frac{s_y}{s_x^2+s_y^2}. \tag{18b}$$

Note that the conventional negative sign in Eq. (18b) is omitted in order to preserve the clockwise/counterclockwise sense of angular measure between mappings. Although the ellipse is distorted in the mapping of Eq. (18), the angles subtended by the slowness ellipse and the extremal distances are all preserved, as is the status of points bounded by the slowness ellipse. Additionally, the orientation of the vector $|s_o\rangle$ is preserved as it is mapped to $|v_o\rangle$. It is this mapping that provides the infrasound analyst with practical uncertainties for direction-of-arrival ($\delta\theta$) and trace velocity (δv).

IV. RESULTS

In this section we apply the method developed in Sec. III to two typical array geometries found in the IMS. The first is the original prototype array recommended by the CTBTO and is used at many of the IMS stations (some arrays will later be augmented with additional sensors beyond the baseline four). Its four elements comprise a regular triangle, 1 km on a side, and its centroid (see Fig. 1, triangles). The second array is a design arrived at by Wilson for use in the IMS stations operated by the University of Alaska Fairbanks. This array is based on two concentric circles, with radii of 1 km and 100 m, respectively. Sensors are placed each at the vertices of a regular pentagon inscribed on the outer circle and, similarly, about a regular triangle inscribed on the inner

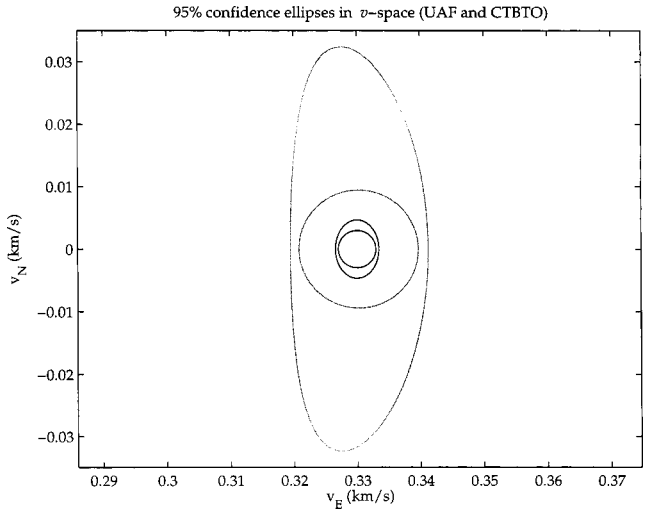


FIG. 2. Slowness confidence ellipsoids mapped to velocity space. Shown are the 95% confidence limits for UAF (black curves) and CTBTO (gray curves) array geometries. For each array type, the inner circles represent a fully operational array; the outer, distorted ellipsoids represent a partially disabled array. Each curve was generated using the following signal parameters: $v_0 = 0.330$ km/s, $\theta_0 = 90°$ and $\sigma_\tau = 0.05$ s.

circle (see Fig. 1, pentagrams). In the field these geometries are often slightly modified, owing to terrain constraints, local land use regulations, or the movement of an ice sheet upon which the sensors are placed.

As the IMS establishes an infrasound station, two salient array performance criteria arise: parameter estimation and robustness to sensor failure. As can be seen from Eq. (5) and Eq. (16), the greater the number of sensors, the less uncertainty in the estimation of signal parameters. Similarly, a greater number of sensors allows for more individual sensor failures before parameter estimation is significantly compromised. A practical concern of the IMS is, however, to keep the cost of deploying a global network of such arrays to a minimum. The trade-off between robustness and parameter estimation may be analyzed by disabling various sensors in the arrays (see Fig. 1, darkened symbols) and calculating the impact on parameter estimation.

A. Event analysis

To ascertain the confidence in parameter estimation from each array type we posit a typical impulsive signal incident upon the array from a direction-of-arrival of $\theta_o = 90°$, moving with a trace velocity of $v_o = 0.330$ km/s. Further, we set $\sigma_\tau = 0.05$ s, as we have found after several years of IMS array operation that $\sigma_\tau \approx 1/f_s$ for impulsive signals; our sampling rate is $f_s = 20$ Hz. In Fig. 2 we depict the results of our uncertainty analysis at the 95% confidence level for each array type, both fully operational and partially disabled. In Table I we summarize the uncertainties defined in Sec. III for the data shown in Fig. 2. It is interesting to note that the fully

TABLE I. Uncertainties calculated for Fig. 2.

	UAF_8	UAF_6	$CTBTO_4$	$CTBTO_3$
$\delta\theta(°)$	1.0	1.6	3.2	11.0
δv (m/s)	5.9	6.9	19.0	22.0

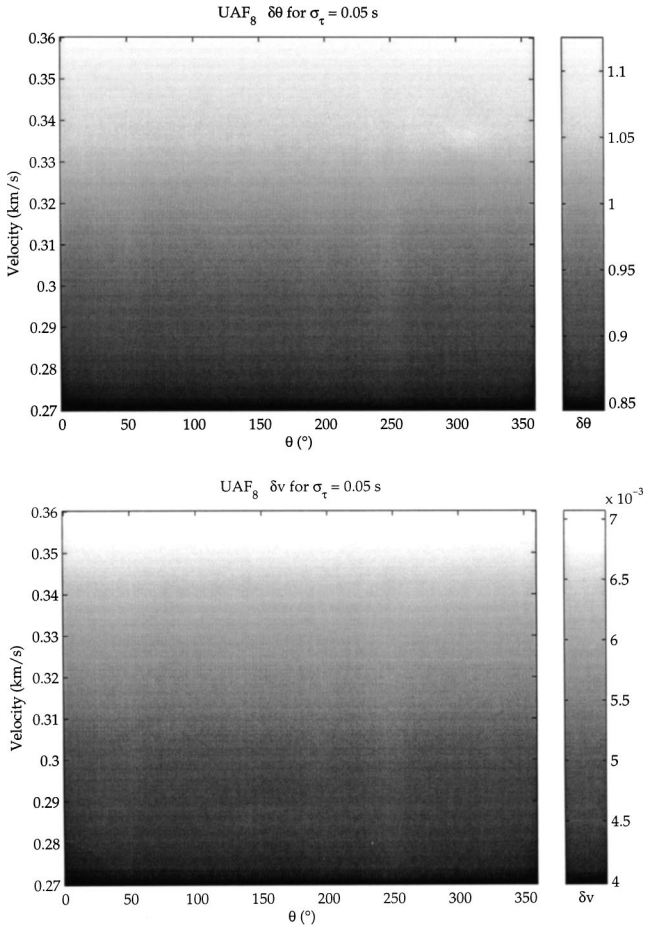


FIG. 3. Array performance bounds for fully operational UAF geometry. Panels depict 95% confidence limits for direction-of-arrival ($\delta\theta$, upper panels) and trace velocity (δv, lower panels) uncertainties as a function of v and θ.

operational four-sensor array cannot match the performance of the eight-sensor array even when two of its sensors have been disabled. Too, the magnitude of performance degradation is much less pronounced with the larger array. While not given in Table I or Fig. 2, we note that in the case of a center sensor failure for the CTBTO array its performance is somewhat degraded ($\delta\theta \approx 3.8°$ and $\delta v \approx 21.8$ m/s), but maintains the circular-symmetric uncertainty properties of the fully operational array.

The ellipsoidal nature of the slowness confidence limit is governed by the dimensionality of the space which the array occupies; therefore, it may have at most a twofold symmetry for a planar array. Confidence in the estimate of the slowness parameter, from which both direction-of-arrival and trace velocity may be derived, is not to be confused with the beam pattern of the array. For example, the beam pattern of the UAF array design would exhibit three- and fivefold symmetries. The ellipsoids of Eq. (17) are distorted by mapping of slowness- to velocity space via Eq. (18). Note that since each array type is circular-symmetric when fully operational, the inner confidence bounds for each array type are circular, unlike their expected beam patterns. In the mapping of Eq. (18), circles map to circles. For a signal incident from due east, the shape of the distorted ellipsoid is consistent with the asymmetry of each disabled array. As the disabled arrays

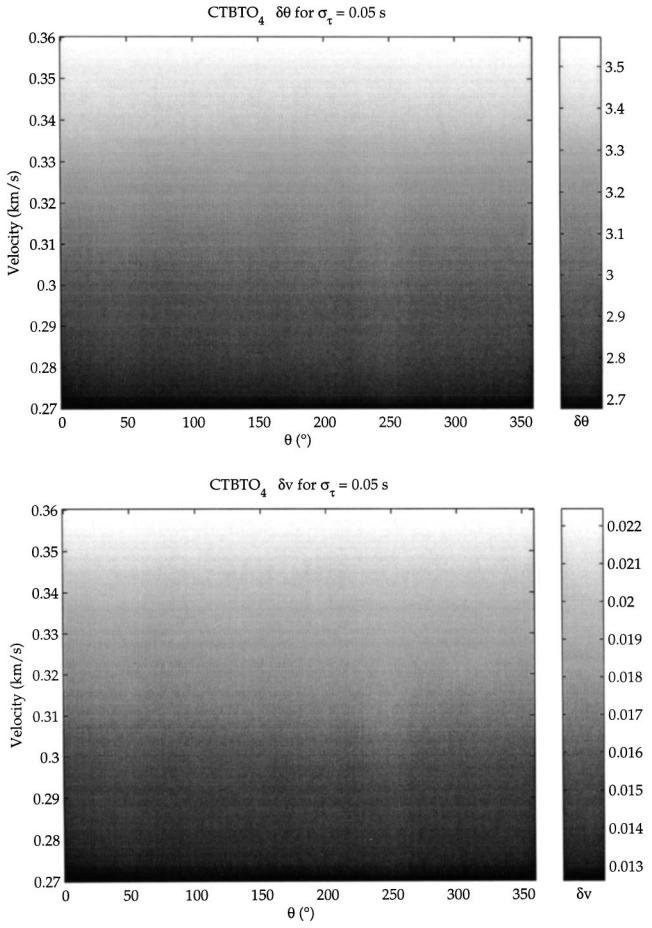


FIG. 4. Array performance bounds for fully operational CTBTO geometry. Panels and parameters are as for Fig. 3.

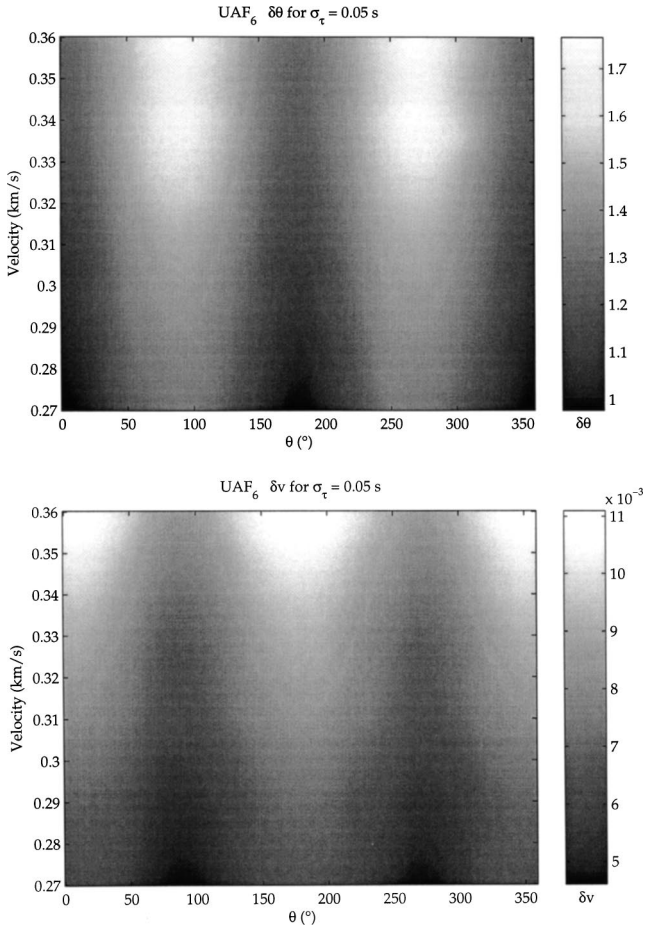


FIG. 5. Array performance bounds for partially disabled (one outer and one inner sensor) UAF geometry. Panels and parameters are as for Fig. 3.

begin to take on an east–west linear form, velocity resolution along the north–south axis is degraded relative to east–west resolution.

B. Array analysis

Whereas particular events may be graphically analyzed after the fashon of Fig. 2, we extend the method of Sec. III to an entire class of signals incident upon an array. To do this we establish a two-dimensional pixel grid of direction-of-arrival and trace velocity coordinates, then perform a calculation at each pixel for $\delta\theta$ and δv, given a confidence limit and σ_τ (95% and 0.05 s for all cases used in this study). The result is displayed as a gray-scale image of the pixel grid, with uncertainty magnitude keyed to gray-scale intensity. We find this technique useful for investigating either the properties of a particular array design or for quantitatively evaluating the performance of one array geometry *vis-à-vis* another. It is the latter use which we present here.

The results of this type of analysis are shown for the fully operational arrays in Figs. 3 and 4, respectively. Similarly, we depict the results for the partially disabled arrays in Figs. 5 and 6. The range of velocities used here is typical of ground-level sound speeds for the various temperatures which might be present at an IMS array. In either parameter, the uncertainty of the four-element CTBTO array exceeds that of the eight-element UAF array by roughly a factor of 3.

Our research group established the pentagram-triangle array design primarily to maximize our detection of certain signal types within the CTBTO passband. An additional concern was the degree of array performance degradation in the event of sensor failure, especially in the case of our IMS array located at Windless Bight, Antarctica, which may only be serviced annually, and at considerable expense. We analyze the impact on parameter uncertainties for the loss of two sensors (one each on the inner and outer rings) on the UAF array (Fig. 5) and one outer sensor on the CTBTO array (Fig. 6). The UAF array suffers parameter estimation confidence degradation by a factor of roughly 1.5, where the CTBTO array degrades by a factor of roughly 3.2.

We note that the uncertainty in either of Figs. 3 and 4 does not vary as a function of direction-of-arrival. This is owing to the circular nature of the confidence ellipsoids for circular-symmetric arrays. Since the sensor losses depicted in Fig. 1 essentially flatten the arrays along an east–west linear form, we expect that the uncertainty in direction-of-arrival will increase along that line; conversely, the uncertainty in trace velocity should decrease. These effects are evident in Figs. 5 and 6. Again, we stress that the uncertainty in parameter estimation is governed by the dimensionality of the array and is quite different from the expected beam pattern for a given geometry.

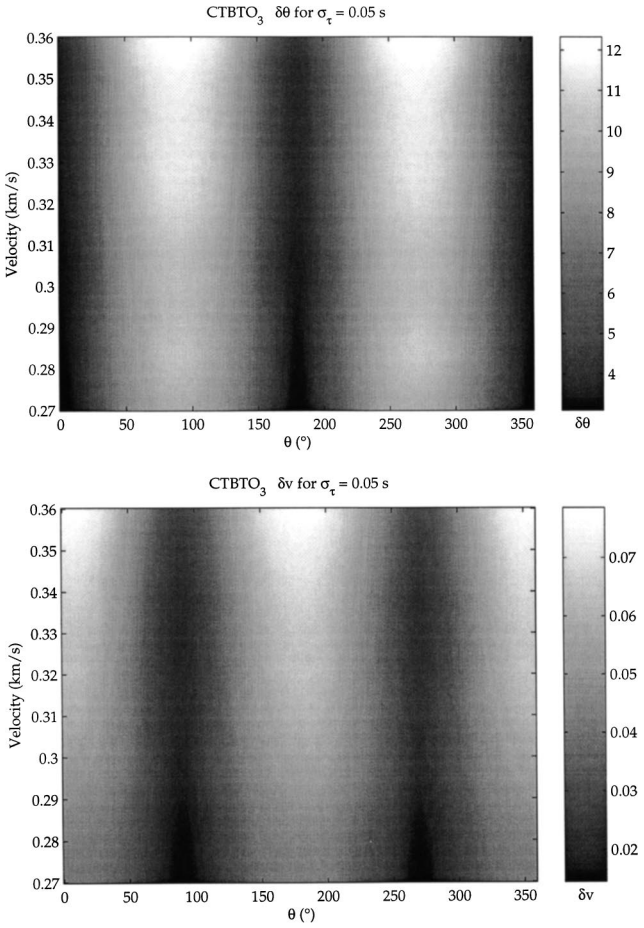


FIG. 6. Array performance bounds for partially disabled (one outer sensor) CTBTO geometry. Panels and parameters are as for Fig. 3.

V. CONCLUSIONS

We have developed a method for determining the uncertainties in parameter estimation for atmospheric infrasound arrays. This method is useful for geometrically describing the uncertainties in direction-of-arrival and trace velocity of a particular event via the confidence ellipsoid determined by Eq. (17). It requires only the estimation of time delays between sensor pairs, following which the uncertainty in those delays may be estimated by Eq. (5). Although we have restricted ourselves to the planar array applications common to the IMS, the method is quite general and may be used in higher dimensional spaces by finding the angle subtended by and the extremal distances to the general ellipsoid given by Eq. (15). Additionally, we have employed the method from an array design standpoint, showing both the relative performance of various array geometries, as well as the impact of sensor loss on parameter estimation. Our research group has found this latter application to be extremely useful, especially in situations where land management policies or topography have prevented us from deploying sensors in optimal locations. In one instance, the method has been employed to monitor the impact on array performance of constant modification to our original array geometry by the flow of an Antarctic ice sheet.

ACKNOWLEDGMENTS

This study was performed as one aspect of the work done evaluating IMS stations for the University of Mississippi. The authors wish to thank Professor C. R. Wilson for many discussions regarding all aspects of array design and deployment. We thank the reviewers for their cogent and insightful comments. One of the authors (C.A.L.S.) thanks the Fairbanks Northstar Borough School District for release time during the course of this study.

[1] Complete information on the Prepatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) and the various monitoring technologies may be found at ⟨http://www.ctbto.org⟩.
[2] B.-G. Song and J. A. Ritcey, ''Angle of arrival estimation of plane waves propagating in random media,'' J. Acoust. Soc. Am. **99**, 1370–1379 (1996).
[3] D. K. Wilson, ''Performance bounds for acoustic direction-of-arrival arrays operating in atmospheric turbulence,'' J. Acoust. Soc. Am. **103**, 1306–1319 (1998).
[4] A. Le Pichon, M. Garcés, E. Blanc, and M. Barthélémy, ''Acoustic propagation and atmosphere characteristics derived from infrasonic waves generated by the Concorde,'' J. Acoust. Soc. Am. **111**, 629–641 (2002).
[5] D. H. Johnson, ''Improving the resolution of bearing in passive sonar arrays by eigenvalue analysis,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-30**, 638–647 (1982).
[6] R. O. Schmidt, ''Multiple emitter location and signal parameter estimation,'' IEEE Trans. Antennas Propag. **AP-34**, 276–280 (1986).
[7] Y. Cansi, ''An automated seismic event processing for detection and location: The P.M.C.C. method,'' Geophys. Res. Lett. **22**, 1021–1024 (1995).
[8] J. L. Spiesberger, ''The matched-lag filter: Detecting broadband multipath signals with auto- and cross-correlation functions,'' J. Acoust. Soc. Am. **109**, 1997–2007 (2002).
[9] B. G. Ferguson, L. G. Criswick, and K. W. Lo, ''Locating far-field impulsive sound sources in air by triangulation,'' J. Acoust. Soc. Am. **111**, 104–116 (2002).
[10] W. L. Donn and B. Naini, ''Sea wave origins of microbaroms and microseisms,'' J. Geophys. Res. **78**, 4482–4488 (1973).
[11] D. Rind, ''Microseisms at Palisades 3. Microseisms and microbaroms,'' J. Geophys. Res. **85**, 4854–4862 (1980).
[12] P. A. M. Dirac, *The Principles of Quantum Mechanics*, 4th ed. (Clarendon, Oxford, 1958).
[13] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice Hall, Upper Saddle River, NJ, 1993).
[14] C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley Series in Probability and Mathematical Statistics (Wiley, New York, 1973).
[15] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing* (Prentice Hall, Englewood Cliffs, NJ, 1992).
[16] W. A. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. (Cambridge University Press, New York, 1992).

Wideband time-reversal imaging of an elastic target in an acoustic waveguide

Lawrence Carin[a] and Hongwei Liu
Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina 27708-0291

Timothy Yoder
SFA Inc., Largo, Maryland 20774

Luise Couchman, Brian Houston, and Joseph Bucaro
Naval Research Laboratory, Physical Acoustics, Code 7130, Washington, District of Columbia 20375-5000

(Received 10 December 2002; revised 25 June 2003; accepted 20 October 2003)

Time-reversal is addressed for imaging elastic targets situated in an acoustic waveguide. It is assumed that the target-sensor range is large relative to the channel depth. We investigate the theory of wideband time-reversal imaging of an extended elastic target, for which the target dimensions are large relative to the principal wavelengths. When performing time-reversal imaging one requires a forward model for propagation through the channel, and the quality of the resulting image may be used as a measure of the match between the modeled and actual (measured) channel parameters. It is demonstrated that the channel parameters associated with a given measurement may be determined via a genetic-algorithm (GA) search in parameter space, employing a cost function based on the time-reversal image quality. Example GA channel-parameter-inversion results and imagery are presented for measured at-sea data. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1632482]

PACS numbers: 43.60.Gk, 43.60.Pt [JJM] Pages: 259–268

I. INTRODUCTION

The use of acoustic backscattered signals to detect and classify underwater targets has attracted significant attention.[1–16] An important issue in sensing underwater targets involves removing the effects of the water channel. Broadhead[14] used a direct method (deterministic deconvolution) to recover the target's scattered signal, although this approach is very sensitive to Green's-function mismatch. Recently, Broadhead, Pflug, and Field[15] developed a more-generalized method by using the minimum entropy deconvolution (MED) method to estimate the channel response function, this requiring that the channel Green's function is sufficiently ''sparse.'' For the case of a very shallow water channel with long observation distances the channel's Green's function is not sparse enough for MED. Sarkissian[16] used an array of sources to approximately generate a low-order mode and an array of receivers to isolate from the received echo a low-order mode, and thus extract the approximately free-field target response in shallow water. In general, this procedure requires a large number of sources, especially for wideband implementation for which many modes are excited.

Kuperman *et al.*[17] and Kim *et al.*[18] have recently demonstrated the time-reversal mirror (Fink[19]) in underwater acoustic environments, with this also referred to as phase conjugation in the frequency domain. Assume a localized source of acoustic energy with the radiated time-domain fields measured by an array of receivers. It has been demonstrated that if the received data is time reversed (phase conjugated) and radiated from the respective sensor locations, focusing is achieved at the source point and the characteristics of the temporal source are approximately recovered (approximately because of the finite array aperture). This phenomenon has significant potential for underwater communications.[17] The same principle may be exploited to remove the channel effects from scattering data and recover the spatial distribution of scattering centers and their temporal dependence. In particular, assume a source generates acoustic energy that subsequently excites a target. When the target scatters the incident radiation it acts as a source. Assume again that the target-radiated (scattered) fields are measured by an array. If the time-domain measured fields are time reversed and radiated from the array into the medium, focusing occurs at the target. This phenomenon has been used experimentally to localize energy on particular points (scattering centers) of a target.[18] In the context of recovering the target scattering characteristics, which we will call ''time-reversal imaging,'' the time-reversed fields are not physically radiated into the medium, but rather this is done numerically. The domain in which the time-reversed fields are computationally radiated constitutes the region in which imaging is effected. This should be contrasted with experimental time reversal interrogation of a target, in which one may perform time reversal repeatedly (''echo mode''), and the system eventually focuses energy on the strongest scatterer or scattering center.[18,20] In numerical time-reversal imaging we recover all scattering centers that radiate energy to the receiver array. The time-reversal image yields a four-dimensional signal (three spatial dimensions and time) for which the time dependence of each scattering center is also

[a]Electronic mail: lcarin@ee.duke.edu

recovered approximately—again, approximately because of the finite receiver array.

In time reversal imaging the three spatial dimensions reveal the location of the scattering centers, and the time dependence characterizes their underlying acoustic and elastic properties. Time-reversal imaging also offers the potential of ''superresolution'' on the target scattering centers. For example, time-reversal imaging of point targets situated in a random media has been investigated previously.[21] It was demonstrated that the multipath caused by the random environment effectively increases the array aperture, yielding improved imaging resolution. A similar superresolution phenomenon[22] is manifested by the multipath in a waveguide, such as an underwater waveguide or channel.[23]

In numerical time-reversal imaging one must computationally propagate the fields through a channel, with this analogous to previous (generally narrow-band) matched-field processing studies.[24] It is demonstrated below that the quality of the time-reversal image is a strong function of the match between the numerical and actual channel parameters. However, this dependence may be exploited to invert for the channel parameters, as has been done in previous matched-field studies.[25–28] In particular, if a strong scatterer is present in the scene under investigation, the focus quality of the associated time-reversal image may be used to quantify the match between the numerical and actual channel parameters. Using a genetic-algorithm[29] for parameter optimization, we demonstrate that one may approximately invert the channel parameters.

The remainder of the text is organized as follows. In Sec. II we present the theory of time-reversal imaging with a particular focus on extended elastic targets and wideband operation. We then address in Sec. III the use of time-reversal imaging techniques to obtain the environmental parameters, in conjunction with a genetic algorithm (GA). The resulting algorithms are then applied in Sec. IV to at-sea experimental data for several elastic targets in an acoustic waveguide. The time-reversal image is coupled with a GA to invert the channel parameters of the channel. In addition, time-reversal-image quality is compared to that of imaging approaches that ignore the channel multipath and elastic nature of the targets. Conclusions are addressed in Sec. V, with recommendations for future research directions.

II. TIME-REVERSAL THEORY FOR EXTENDED ELASTIC TARGETS

A. Time-reversal imaging of locally reactive targets

Consider acoustic scattering from an elastic target with surface Γ (see Fig. 1). The target is insonified by the acoustic field produced by a volume source distribution, P_i, and the acoustic signal scattered from the target and measured at $\boldsymbol{r}_r$ is $P_s(\boldsymbol{r}_r)$. A linear operator representation of this problem may be expressed as

$$P_i \rightarrow \underline{A} \rightarrow P_{isur}(\boldsymbol{r}'') \rightarrow \underline{B} \rightarrow P_{tsur}(\boldsymbol{r}') \rightarrow \underline{S} \rightarrow P_s(\boldsymbol{r}_r). \tag{1}$$

The operator $\underline{A}$ maps P_i to $P_{isur}(\boldsymbol{r}'')$, representative of the incident pressure (if the target were absent) at $\boldsymbol{r}''$ on the target surface. The operator $\underline{B}$ maps $P_{isur}(\boldsymbol{r}'')$ to $P_{tsur}(\boldsymbol{r}')$, the total pressure on the target surface at $\boldsymbol{r}'$. Finally $\underline{S}$ is the operator that maps $P_{tsur}(\boldsymbol{r}')$ to $P_s(\boldsymbol{r}_r)$.

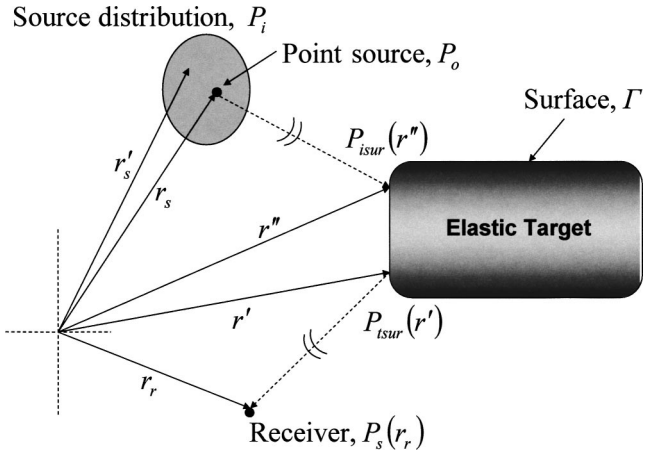


FIG. 1. Schematic of coordinate system used to describe time-reversal imaging.

The operator $\underline{A}$ may be expressed as

$$P_{isur}(\boldsymbol{r}'') = \int_{v'} P_i(\boldsymbol{r}_s') G(\boldsymbol{r}'', \boldsymbol{r}_s') dV', \tag{2}$$

where $G(\boldsymbol{r}'', \boldsymbol{r}_s')$ is the Green's function for the acoustic medium. We simplify the discussion by considering a single point source with amplitude, P_o, at location $\boldsymbol{r}_s$, i.e., $P_i(\boldsymbol{r}_s') = P_o \delta(\boldsymbol{r}_s' - \boldsymbol{r}_s)$. Given a point source, $\underline{A}$ becomes:

$$P_{isur}(\boldsymbol{r}'') = P_o G(\boldsymbol{r}'', \boldsymbol{r}_s). \tag{3}$$

The relationship $\underline{B}$ between $P_{isur}(\boldsymbol{r}'')$ and $P_{tsur}(\boldsymbol{r}')$ may be expressed as a general linear operator

$$P_{tsur}(\boldsymbol{r}') = \int_{\Gamma} P_{isur}(\boldsymbol{r}'') X(\boldsymbol{r}', \boldsymbol{r}'') d\Gamma'' \tag{4}$$

where $X(\boldsymbol{r}', \boldsymbol{r}'')$ represents the ''transfer function'' that maps the incident field on the elastic body to the total field on the body. The operator $X(\boldsymbol{r}', \boldsymbol{r}'')$ generally includes the rigid and elastic components of scattering. Finally, the operator $\underline{S}$ is the Helmholtz integral[24]

$$P_s(\boldsymbol{r}_r) = \int_{\Gamma} \left[P_{tsur}(\boldsymbol{r}') \frac{\partial G(\boldsymbol{r}, \boldsymbol{r}')}{\partial n'} - G(\boldsymbol{r}_r, \boldsymbol{r}') \frac{\partial P_{tsur}(\boldsymbol{r}')}{\partial n'} \right] d\Gamma'. \tag{5}$$

Substituting (3) and (4) into (5) we arrive at

$$P_s(\boldsymbol{r}_r, \boldsymbol{r}_s) = P_o \int_{\Gamma} \int_{\Gamma} \left[G(\boldsymbol{r}'', \boldsymbol{r}_s) X(\boldsymbol{r}', \boldsymbol{r}'') \frac{\partial G(\boldsymbol{r}_r, \boldsymbol{r}')}{\partial n'} - G(\boldsymbol{r}_r, \boldsymbol{r}') G(\boldsymbol{r}'', \boldsymbol{r}_s) \frac{\partial X(\boldsymbol{r}', \boldsymbol{r}'')}{\partial n'} \right] d\Gamma'' d\Gamma'. \tag{6}$$

We initially simplify the discussion by first considering an infinite homogenous medium characterized by wavenumber β, and therefore, $(\nabla^2 + \beta^2) G(\boldsymbol{r}_r, \boldsymbol{r}') = -\delta(\boldsymbol{r}_r - \boldsymbol{r}')$ and $G(\boldsymbol{r}_r, \boldsymbol{r}') = \exp(-j\beta|\boldsymbol{r}_r - \boldsymbol{r}'|)/4\pi|\boldsymbol{r}_r - \boldsymbol{r}'|$, where we have assumed and suppressed an $\exp(j\omega t)$ time dependence.

Since $\partial G(\boldsymbol{r}_r,\boldsymbol{r}')/\partial n' = -(j\beta + 1/\|\boldsymbol{r}_r-\boldsymbol{r}'\|)G(\boldsymbol{r}_r,\boldsymbol{r}')(\partial\|\boldsymbol{r}_r-\boldsymbol{r}'\|/\partial n')$, Eq. (6) becomes

$$P_s(\boldsymbol{r}_r,\boldsymbol{r}_s) = P_o \int_\Gamma \int_\Gamma G(\boldsymbol{r}'',\boldsymbol{r}_s)u(\boldsymbol{r}_s,\boldsymbol{r}'',\boldsymbol{r}',\boldsymbol{r}_r) \times G(\boldsymbol{r}_r,\boldsymbol{r}')d\Gamma''d\Gamma', \quad (7)$$

where

$$u(\boldsymbol{r}_s,\boldsymbol{r}'',\boldsymbol{r}',\boldsymbol{r}_r) = X(\boldsymbol{r}',\boldsymbol{r}'')\left(j\beta + \frac{1}{\|\boldsymbol{r}_r-\boldsymbol{r}'\|}\right)\frac{\partial\|\boldsymbol{r}_r-\boldsymbol{r}'\|}{\partial n'} + \frac{\partial X(\boldsymbol{r}',\boldsymbol{r}'')}{\partial n'}, \quad (8)$$

represents a ''transfer function'' from the unperturbed (no target present) incident pressure at the target surface to a virtual source distribution (no target present) that would produce the scattered field at $\boldsymbol{r}_r$. We note that $u(\boldsymbol{r}_s,\boldsymbol{r}'',\boldsymbol{r}',\boldsymbol{r}_r)$ accounts for all interactions on the target. Assuming that the target is locally reactive, $u(\boldsymbol{r}_s,\boldsymbol{r}'',\boldsymbol{r}',\boldsymbol{r}_r)=0$ if $\boldsymbol{r}''\neq\boldsymbol{r}'$, and (7) becomes

$$P_s(\boldsymbol{r}_r,\boldsymbol{r}_s) = P_o \int_\Gamma G(\boldsymbol{r}',\boldsymbol{r}_s)u(\boldsymbol{r}_s,\boldsymbol{r}',\boldsymbol{r}_r)G(\boldsymbol{r}_r,\boldsymbol{r}')d\Gamma'. \quad (9)$$

By ''locally reactive'' we are here assuming no interactions between the scattering centers on the target, and hence that each scattering center acts in isolation (i.e., locally).

We now assume that $P_o=1$ and that the target is large enough with respect to wavelength that a stationary-phase analysis[30] of the integral in (9) is appropriate. Assuming M stationary scattering points, we have

$$P_s(\boldsymbol{r}_r,\boldsymbol{r}_s) \approx \sum_{m=1}^{M(\boldsymbol{r}_r,\boldsymbol{r}_s)} G(\boldsymbol{r}'_m,\boldsymbol{r}_s)\hat{u}(\boldsymbol{r}_s,\boldsymbol{r}'_m,\boldsymbol{r}_r)G(\boldsymbol{r}_r,\boldsymbol{r}'_m), \quad (10)$$

where $\hat{u}(\boldsymbol{r}_s,\boldsymbol{r}'_m,\boldsymbol{r}_r)$ accounts for the new factors associated with the stationary-phase analysis.[30] Note that the form in (10) is analogous to that considered in Ref. 21 for scattering from M point scatterers; the distinction here is that $\hat{u}(\boldsymbol{r}_s,\boldsymbol{r}'_m,\boldsymbol{r}_r)$ accounts for an *extended* target on which there are M scattering centers of importance. The number of significant scattering centers is a function of the source and observation points, as indicated in (10). In Sec. II C we remove the ''locally reactive'' simplification and, therefore, account for all interactions (multiple scattering) between the M scattering centers.

Assume now a single source located at $\boldsymbol{r}_s$, and a linear array of receivers (see Fig. 2), the kth receiver located at $\boldsymbol{r}_k$. Moreover, assume that the array aperture (size) is such that the same M scattering centers are observed at all $\boldsymbol{r}_k$. Explicitly employing the form of the Green's function and assuming that the target-sensor distance R is large relative to the size of the target and aperture, (10) becomes for the kth receiver

$$P_s(\boldsymbol{r}_k;\boldsymbol{r}_s) \approx \sum_{m=1}^{M} \hat{u}_m(\boldsymbol{r}_k,\boldsymbol{r}_s,\omega)\exp(-j\beta|\boldsymbol{r}'_m-\boldsymbol{r}_s|) \times \exp(-j\beta|\boldsymbol{r}'_m-\boldsymbol{r}_k|), \quad (11)$$

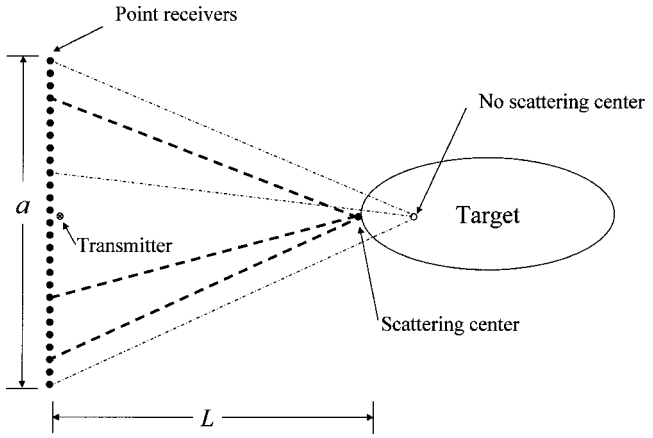


FIG. 2. Schematic of array-based sensing of an extended target. A single source is assumed at the center of the vertical receiver array. A scattering center is assumed at the front of the target, and focusing of the time-reversal image occurs here, since the delays $\Omega_{im}(\boldsymbol{r};\boldsymbol{r}'_m,\boldsymbol{r}_s,\boldsymbol{r}_k)$ in (12) are identical at the scattered center $\boldsymbol{r}=\boldsymbol{r}'_m$. Due to the varying time delays, at positions $\boldsymbol{r}$ away from the scattering center, generally destructive interference is manifested.

where $\hat{u}_m(\boldsymbol{r}_k,\boldsymbol{r}_s,\omega)\equiv(1/R^2)\hat{u}(\boldsymbol{r}_k,\boldsymbol{r}'_m,\boldsymbol{r}_s,\omega)$ and the frequency, ω, dependence is explicitly stated. The term $\hat{u}_m(\boldsymbol{r}_k,\boldsymbol{r}_s,\omega)$ accounts for radiation emitted to the kth receiver from the mth scattering center. The expression in (11) represents the measured signal due to a source at $\boldsymbol{r}_s$ and a receiver at $\boldsymbol{r}_k$. Using multiple receivers (bistatic measurement) our goal is to image the M scattering centers. We perform phase conjugation of (11) in the frequency domain and then propagate this signal numerically into the computational domain from the receiver at $\boldsymbol{r}_k$. This corresponds to time reversal[19] in the time domain. This time reversal process accounts for propagation from the scattering centers to the receiver. To account for propagation from the source to the scattering center we also multiply the aforementioned phase-conjugated response by the Green's function characteristic of propagation from the source to the scattering center. For the free-space problem discussed above such propagation from the source and receiver may be implemented in closed form via the available Green's function. For more-complex media such computations must be performed numerically.[24]

Implementing the above phase conjugation, accounting for propagation from the source $\boldsymbol{r}_s$ and the receiver $\boldsymbol{r}_k$ to point $\boldsymbol{r}$ in the image domain, and finally performing an inverse Fourier transform to convert the signal to the time domain, we obtain the space–time signal

$$I_{ks}(\boldsymbol{r},t) = \sum_{m=1}^{M} \int \hat{u}_m^*(\boldsymbol{r}_k,\boldsymbol{r}_s,\omega) \times \exp\{j\omega[t-\Omega_m(\boldsymbol{r};\boldsymbol{r}'_m,\boldsymbol{r}_s,\boldsymbol{r}_k)]\}d\omega, \quad (12)$$

where $\Omega_m(\boldsymbol{r};\boldsymbol{r}'_m,\boldsymbol{r}_s,\boldsymbol{r}_k)=\beta\{-|\boldsymbol{r}'_m-\boldsymbol{r}_s|-|\boldsymbol{r}'_m-\boldsymbol{r}_k|+|\boldsymbol{r}-\boldsymbol{r}_s|+|\boldsymbol{r}-\boldsymbol{r}_k|\}$, and superscript* represents complex conjugation. The expression in (12) is simply a time-shifted version of the same time-dependent signal for each observation point $\boldsymbol{r}$. The cumulative time-reversal image for a source at $\boldsymbol{r}_s$ is obtained by summing over all K receivers, $I_s(\boldsymbol{r},t)=\Sigma_{k=1}^{K} I_{ks}(\boldsymbol{r},t)$. If we assume that the virtual sources associated with each scattering center are approximately the same, as observed at each

observation point [i.e., $\hat{u}_m(\boldsymbol{r}_k,\boldsymbol{r}_s,\omega)\approx\hat{u}_m(\boldsymbol{r}_s,\omega)$], then within $I_s(\boldsymbol{r},t)$ the time dependence of the scattering coefficients add coherently (constructively) at the scattering centers $\boldsymbol{r}=\boldsymbol{r}'_m$ with incoherent (destructive) interference elsewhere, yielding the final space–time image

$$I_s(\boldsymbol{r},t)=\sum_{k=1}^{K}\sum_{m=1}^{M}\int \hat{u}^*_m(\boldsymbol{r}_s,\omega)\times\exp\{j\omega[t-\Omega_m(\boldsymbol{r};\boldsymbol{r}'_m,\boldsymbol{r}_s,\boldsymbol{r}_k)]\}d\omega. \quad (13)$$

We assume the array of receivers has inter-element separation less than one-half wavelength at the highest frequency of interest thereby avoiding aliasing or equivalently sidelobes.[31] Because we have accounted in (13) for propagation from the source to the scattering center, the time reference *of each pixel* in the image, $t=0$, corresponds to the arrival of the associated incident wave.

B. Time-reversal imaging resolution

In Fig. 2 we consider a linear aperture and address imaging of one of the target's scattering centers. For simplicity the example is considered in two dimensions. When the imaging point $\boldsymbol{r}$ is situated at the scattering center ($\boldsymbol{r}=\boldsymbol{r}'_m$), the shift $\Omega_m(\boldsymbol{r};\boldsymbol{r}'_m,\boldsymbol{r}_s,\boldsymbol{r}_k)=0$ for all receivers k and $I_s(\boldsymbol{r},t)=K\Sigma_{m=1}^{M}\int\hat{u}^*_m(\boldsymbol{r}_s,\omega)\exp(j\omega t)d\omega$, yielding coherent addition of the responses from the K receiver positions. As $\boldsymbol{r}$ moves away from the scattering center (i.e., with increasing $|\boldsymbol{r}-\boldsymbol{r}'_m|$), the $\Omega_m(\boldsymbol{r};\boldsymbol{r}'_m,\boldsymbol{r}_s,\boldsymbol{r}_k)\neq 0$ becomes different for distinct k, and the signals interfere destructively when we sum over index k to form $I_s(\boldsymbol{r},t)$. From the geometrical construct presented in Fig. 2, we note that the differences in $\Omega_m(\boldsymbol{r};\boldsymbol{r}'_m,\boldsymbol{r}_s,\boldsymbol{r}_k)$ are more pronounced as a function of k as the aperture a increases and as L decreases. In addition, as the frequency of the signal increases (reduced wavelength), the signals $\Sigma_{m=1}^{M}\int\hat{u}^*_m(\boldsymbol{r}_s,\omega)\exp\{j\omega[t-\Omega_m(\boldsymbol{r};\boldsymbol{r}'_m,\boldsymbol{r}_s,\boldsymbol{r}_k)]\}d\omega$ change more quickly with time, enhancing the defocusing for given k-dependent temporal shifts $\Omega_m(\boldsymbol{r}\neq\boldsymbol{r}_m;\boldsymbol{r}'_m,\boldsymbol{r}_s,\boldsymbol{r}_k)$. Therefore, it is anticipated that the focus tightness is inversely proportional to a, directly proportional to L, and directly proportional to the wavelength λ (for narrowband operation). In particular, the minimum focus is $\lambda L/a$ parallel to the linear array and $\lambda(L/a)^2$ in the orthogonal direction.[31] A detailed study of time-reversal resolution is presented in Ref. 21 (Papanicolaou), wherein the above resolution terms have been derived explicitly. Moreover, in Ref. 22 it is demonstrated that multi-path may effectively increase the aperture size a, yielding ''super-resolution'' in time-reversal imaging.

C. Multiple sources and time-reversal imaging of nonlocally reactive targets

Two assumptions have been made in the analysis presented above. First, we have assumed a single source located at $\boldsymbol{r}_s$. If we have N_s sources the total time-reversal image is $I(\boldsymbol{r},t)=\Sigma_{s=1}^{N_s}I_s(\boldsymbol{r},t)$. The fields scattered from a given scattering center are generally excitation dependent, and therefore, each of the N_s sources will in general produce a unique contribution to the time-dependent time-reversal signal at a given scattering center. For example, some scattering centers may only be excited strongly by a subset of sources.

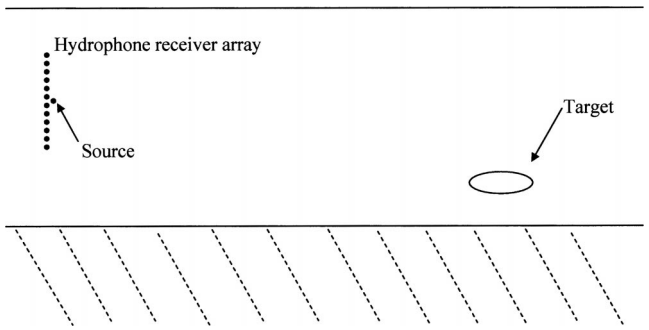


FIG. 3. Schematic of a target in an acoustic waveguide. A single isotropic source is employed, with a vertical array of receivers. The position of the array and source are kept at a constant depth and moved along a line, to constitute a linear synthetic aperture.

In Sec. III we consider sensing with a linear vertical array of receivers in an acoustic waveguide excited by a single source, with waveguide depth small compared to the ranges of interest (see Fig. 3). Although this problem only employs a single *physical* source, the multiple reflections inside the waveguide yield a set of *virtual* image sources constituting an effective multi-source problem. It is well known that such a problem may be parameterized in terms of an infinite number of vertical sources (image sources) situated in free field, with each source weighted by an appropriate number of reflection coefficients at the top and bottom of the waveguide.[23] A similar phenomenon is associated with the vertical receiver array. The set of virtual image sources yield an effective aperture a that is larger than the physical aperture, providing an example for which the aforementioned super-resolution phenomenon is realized.

The second simplification in Sec. II A is the assumption of a locally reactive target, as discussed prior to Eq. (9). Many targets are not locally reactive, and therefore we reconsider $u(\boldsymbol{r}_s,\boldsymbol{r}'',\boldsymbol{r}',\boldsymbol{r}_r)$. As in (10), assume that the fields scatter to the receivers from M principal scattering centers $\boldsymbol{r}'_m$. We now also assume that energy interacts between points $\boldsymbol{r}''_p$ and $\boldsymbol{r}'_m$ on the target surface. If we assume N_{pm} interactions between these points then $u(\boldsymbol{r}_s,\boldsymbol{r}''_p,\boldsymbol{r}'_m,\boldsymbol{r}_r)\approx u_m(\boldsymbol{r}_s,\boldsymbol{r}_r)\Sigma_{n=0}^{N_{pm}}\gamma_{n,pm}(\omega)\exp(-j\beta_{pm}2n|\boldsymbol{r}'_m-\boldsymbol{r}''_p|)$, where β_{pm} accounts for the wave speed between the two scattering centers and the $n=0$ term represents direct diffraction from the mth scattering center to the receiver. Equation (11) now generalizes as

$$P_s(\boldsymbol{r}_k;\boldsymbol{r}_s)\approx\sum_{p=1}^{P}\sum_{m=1}^{M}\hat{u}_m(\boldsymbol{r}_s,\omega)\exp(-j\beta|\boldsymbol{r}'_m-\boldsymbol{r}_s|)\times\exp(-j\beta|\boldsymbol{r}'_m-\boldsymbol{r}_k|)\sum_{n=0}^{N_{mp}}\gamma_{n,pm}\times\exp(-j\beta_{pm}2n|\boldsymbol{r}'_m-\boldsymbol{r}^n_p|). \quad (14)$$

If we now produce a time-reversal image as in (13), we have

$$I_s(\mathbf{r},t)=\sum_{k=1}^{K}\sum_{p=1}^{P}\sum_{m=1}^{M}\sum_{n=0}^{N_{mp}}\int \hat{u}_m^*(\mathbf{r}_s,\omega)\gamma_{n,pm}(\omega)$$
$$\times\exp\{j\omega[t-\Omega_{mn}(\mathbf{r};\mathbf{r}'_m,\mathbf{r}''_p,\mathbf{r}_s,\mathbf{r}_k)]\}d\omega, \quad (15)$$

where $\Omega_{mn}(\mathbf{r};\mathbf{r}'_m,\mathbf{r}''_p,\mathbf{r}_s,\mathbf{r}_k)=-\beta_{pm}2n|\mathbf{r}'_m-\mathbf{r}''_p|+\beta\{-|\mathbf{r}'_m-\mathbf{r}_s|-|\mathbf{r}'_m-\mathbf{r}_k|+|\mathbf{r}-\mathbf{r}_s|+|\mathbf{r}-\mathbf{r}_k|\}$. When $\mathbf{r}=\mathbf{r}'_m$ we again obtain focusing, since $\beta\{-|\mathbf{r}'_m-\mathbf{r}_s|-|\mathbf{r}'_m-\mathbf{r}_k|+|\mathbf{r}-\mathbf{r}_s|+|\mathbf{r}-\mathbf{r}_k|\}=0$ for all source and receiver points, $\mathbf{r}_s$ and $\mathbf{r}_k$. However, now the signal imaged at $\mathbf{r}=\mathbf{r}'_m$ also accounts for time-delayed responses due to interactions between scattering centers m and p. The time delays are realized by the phase $\beta_{pm}2n|\mathbf{r}'_m-\mathbf{r}''_p|$, this accounting for interactions between scattering centers m and p and, therefore, removing the ''locally reactive'' assumption. Note that $\beta_{pm}2n|\mathbf{r}'_m-\mathbf{r}''_p|$ is again the same for all observation points.

III. CHANNEL PARAMETER INVERSION VIA TIME-REVERSAL IMAGING

In the above discussion we considered a target in free field (no channel) for which the Green's function is known exactly. For the case of a source in a waveguide (acoustic channel) with a vertical array of receivers the Green's function is a function of many channel parameters including the water sound speed, the channel depth, and the properties of the bottom. Assume that G_a represents the actual Green's function of the channel considered in a given measurement and that G_{TR} represents the Green's function employed numerically when performing time-reversal imaging. Then, as extension of (9) and (10), the measured signals are represented as

$$P(\mathbf{r}_k;\mathbf{r}_s)\approx\sum_{m=1}^{M(\mathbf{r}_k,\mathbf{r}_s)}\hat{u}(\mathbf{r}_k,\mathbf{r}'_m,\mathbf{r}_s)G_a(\mathbf{r}'_m,\mathbf{r}_s)G_a(\mathbf{r}_k,\mathbf{r}'_m), \quad (16)$$

and the space–time time-reversal image is expressed as

$$I_s(\mathbf{r},t)=\sum_{k=1}^{K}\int P^*(\mathbf{r}_k;\mathbf{r}_s)G_{\mathrm{TR}}(\mathbf{r},\mathbf{r}_s)G_{\mathrm{TR}}(\mathbf{r}_k,\mathbf{r})$$
$$\times\exp(i\omega t)d\omega. \quad (17)$$

For simplicity we consider locally reactive targets, recognizing that the more general target case is a straightforward extension using descriptions in Sec. II C. When performing time-reversal imaging we must estimate the appropriate Green's function G_{TR}. Specifically, both G_a and G_{TR} represent waveguide Green's functions, but it is important that the channel parameters associated with G_{TR} are consistent with the actual channel parameters (embodied in G_a).

We next employ the time-reversal image itself to estimate the appropriate parameters for G_{TR}, assuming that a strong scatterer or scattering center is present in the domain being imaged. In particular, assume channel parameters Φ are employed in G_{TR} and that the associated time-reversal image is represented as $I_s(\mathbf{r},t;\Phi)$. We treat the inversion for the channel parameters as an optimization problem and find the set of parameters that minimize a cost function defined in terms of $I_s(\mathbf{r},t;\Phi)$. Unfortunately, the parameters Φ may be of high dimension, and therefore, the search space may be so large that an exhaustive search for the optimal parameters is not generally possible. In addition, the dependence of the cost function on the search-space parameters may be highly discontinuous, with local minima, precluding the use of calculus-based optimization methods (gradient-search methods). Genetic algorithms (GAs) and simulated annealing (SA) are two general techniques for multi-dimensional optimization problems, and these have been used for acoustic-channel inversion.[25–28] In the examples to follow, when generating the space–time image $I_s(\mathbf{r},t;\Phi)$, the waveguide Green's function $G_{\mathrm{TR}}(\Phi)$ is implemented via a normal-mode model.[24]

The time-reversal image $I_s(\mathbf{r},t;\Phi)$ is a four-dimensional quantity (space–time). For channel inversion we simplify it as follows. The space-dependent image is defined as

$$\hat{I}_s(i,j,k;\Phi)\equiv I_s(i,j,k,t=0;\Phi), \quad (18)$$

where (i,j,k) denote the sampled points in three-dimensional space, and we observe the signal at time $t=0$ [from (13)] which corresponds to integrating the frequency-domain signal for all frequencies. Recall that, because the time-reversal imaging has accounted for propagation from the source to the target and from the target to the receiver, time $t=0$ corresponds to the time at which specular reflection (wavefront scattering) occurs at *each* scattering center. Hence, the time $t=0$ corresponds to the time-reversal image with largest energy, with times $t>0$ corresponding to weaker resonant scattering (multiple scattering between scattering centers). For a *single* vertical array we have no resolution in cross range, and we denote the depth and down-range spatial dimensions (i,j) yielding $\hat{I}_s(i,j;\Phi)$. As discussed further below, the cross-range resolution is achieved via a synthetic aperture of multiple vertical arrays with data collected along a (synthetic) linear aperture.

Assuming that a strong scatterer exists in the vicinity of the region $(i,j)\in\Lambda$, we define the equivalent probability mass function (pmf)

$$\mathrm{p}(i,j;\Phi)=\frac{|\hat{I}_s(i,j;\Phi)|}{\Sigma_{(i,j)\in\Lambda}|\hat{I}_s(i,j;\Phi)|}, \quad (19)$$

where $\mathrm{p}(i,j;\Phi)$ is defined for $(i,j)\in\Lambda$. Equation (19) is motivated by the fact that a pmf should be positive, and hence, we use the magnitude of $\hat{I}_s(i,j;\Phi)$; moreover, the normalization in Eq. (19) is dictated by the fact that the pmf should sum to one.

If the image is tightly focused about the strong scatterer located in the region Λ, the $\mathrm{p}(i,j;\Phi)$ should have low entropy.[32] By contrast, the entropy increases as the image defocuses, and once $\mathrm{p}(i,j;\Phi)$ has uniform weighting for all (i,j) corresponding to maximum defocusing, the entropy is maximum. The first component of the GA cost function is the entropy of $\mathrm{p}(i,j;\Phi)$, defined as[32]

$$\mathrm{E}_1(\Phi)=-\sum_{(i,j)\in\Lambda}\mathrm{p}(i,j;\Phi)\ln\mathrm{p}(i,j;\Phi). \quad (20)$$

The second component of the cost function is analogous to that discussed in Refs. 25–28. Let $r_k(\omega_i)$ represent the scattered signal measured at receiver k for frequency ω_i.

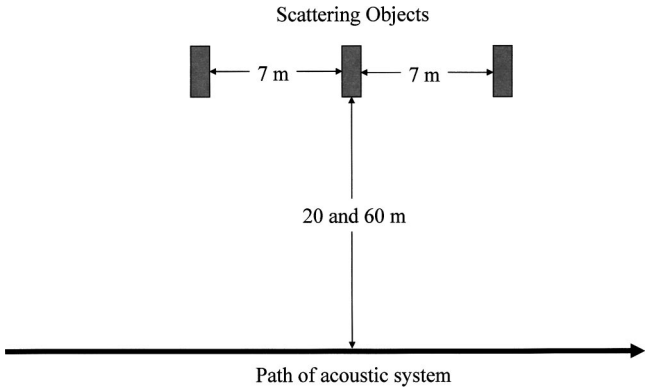


FIG. 4. Schematic of experimental geometry.

Further, let $c_k(\omega_i;i,j,\Phi)$ represent the *computed* scattered response at frequency ω_i for a point scatterer located at a point $(i,j)\in\Lambda$. Assuming N_f frequencies the second component to the cost function is

$$\mathrm{E}_2(\Phi) = 1 - \max_{(i,j)\in\Lambda} \times \left(\frac{1}{N_f} \sum_{n=1}^{N_f} \frac{|\Sigma_{k=1}^{K} r_k(\omega_n) c_k(\omega_i;i,j,\Phi)|^2}{\Sigma_{k=1}^{K} |r_k(\omega_n)|^2 \Sigma_{k=1}^{K} |c_k(\omega_i;i,j,\Phi)|^2} \right), \tag{21}$$

and the total cost function is $\mathrm{E}(\Phi)=\mathrm{E}_1(\Phi)+\gamma\mathrm{E}_2(\Phi)$. Equation (21) is a measure of the correlation across frequency between the measured and computed scattered waveforms, offset such that a high level of correlation implies E_2 near zero. A genetic algorithm (GA) is employed to determine those parameters Φ that minimize $\mathrm{E}(\Phi)$, and we have found that setting $\gamma=1$ yields good results. In the experiments presented in Sec. IV, the approximate location of the strong scatterer is assumed known, and this defines Λ.

A comment is in order concerning the combined cost function $\mathrm{E}(\Phi)=\mathrm{E}_1(\Phi)+\gamma\mathrm{E}_2(\Phi)$ and its individual components $\mathrm{E}_1(\Phi)$ and $\mathrm{E}_2(\Phi)$. Through extensive numerical studies we have found that GA optimization of E_1 alone will converge to the proper parameters, but more slowly than for the composite cost function E. In addition, E_2 alone was found to have numerical difficulties with local minima. By combining these two components, we have consistently found good convergence properties of E across a large range of measured and computed data.

IV. TIME-REVERSAL IMAGING WITH MEASURED SEA DATA

A. At-sea experimental data

The experimental data used to demonstrate the time-reversal and channel parameter inversion techniques were collected by the Naval Research Laboratory (Washington, DC) and Coastal Systems Station (Panama City). Three targets were placed on the bottom of St. Andrews bay (near Panama City, FL, USA) in water that was approximately 8 m deep. These targets were approximately 1 to 2 m long, 0.5 m wide, and 0.5 m tall. The targets were spaced approximately 7 m apart and oriented as depicted in Fig. 4. The bottom of the bay is relatively flat over the 60 m range from the target field to the acoustic system that collected the data. The channel parameters measured at the experimental location are listed in Table I. The acoustic system consisted of a vertical array of 12 sensors spaced 7.62 cm apart, with an isotropic source situated at the bottom of the array. The middle of the array was 3.4 m below the air–water interface, and the sensor bandwidth was 1.3–10 KHz. The experimental data that we use were collected for multiple positions of the vertical array, along a linear path (see Fig. 4). The spacing between data collection points along this linear path was approximately 0.22 m.

B. Channel parameter inversion

Here we present results for channel-parameter inversion using the measured data discussed above. We estimate the channel parameters employing a GA and the cost function (discussed in Sec. III) operating on the measured acoustic signals scattered from the bottom-mounted targets. The goal of this section is to demonstrate that the quality of the time-reversal image may be used on measured at-sea data to approximately invert for the channel parameters. Moreover, after performing such GA-based channel-parameter inversion, in Sec. IV C we use the derived channel parameters to form time-reversal synthetic-aperture-sonar (SAS) images of the targets in question, and compare these to conventional SAS imaging. The overall objective, therefore, is to demonstrate on measured data that inversion for the channel parameters is feasible using time-reversal imaging, and that the channel parameters so determined may then be used to generate enhanced time-reversal SAS imagery of three-dimensional targets.

The channel parameters, Φ, considered were the water depth, water sound speed, sound speed in the bottom, the density of the bottom, the attenuation of the bottom, target range, and target depth (a total of seven parameters). For cost-function component E_1 we must perform time-reversal imaging over (i,j) for each set of channel parameters with the same true for E_2. It is therefore necessary that a fast forward model be employed, and here the simplicity of the wave-

TABLE I. Channel parameters estimated during the measurement, estimated via the GA optimization, and the parameter search range employed by the GA.

	Measured value	Estimated value	Search range
Water depth(m)	8.23	8.3	(7.73–8.73)
SS in water(m/s)	1490	1489.21	(1440–1540)
SS in sediment(m/s)	1705	1655.25	(1605–1805)
Density of sediment(g/cm 3)	N/A	3.086	(1–3.5)
Attenuation of sediment(dB/kmHz)	N/A	0.504	(0.1–0.9)
Target distance (m)	60	60.73	(55–65)
Target depth(m) (to top of target)	7.70	7.744	(7–8.73)

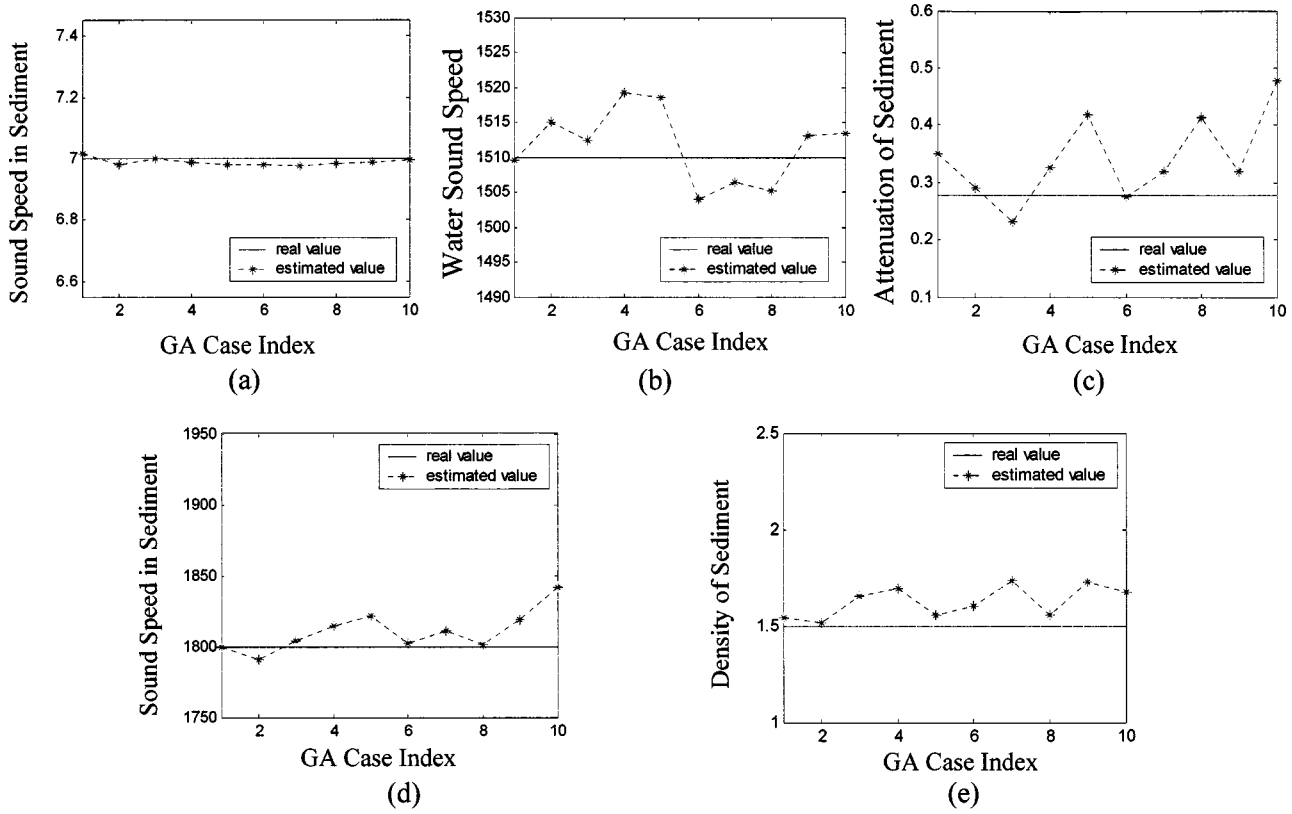


FIG. 5. Channel parameters determined by the genetic-algorithm optimization, for ten different, randomly chosen initial GA channel-parameter-space populations. (a) channel depth, (b) sound speed in the channel, (c) attenuation in the bottom, (d) sound speed in the bottom, (e) density of the bottom.

guide (constant or near-constant channel depth) allowed use of a modal model.[24]

For the results presented below each GA population consisted of 30 sets of parameters, Φ, and the algorithm converged after 50 GA iterations. A conventional GA algorithm was applied.[29] Briefly, the GA employed the cost function $E(\Phi)=E_1(\Phi)+\gamma E_2(\Phi)$ to rank the "fitness" of the 30 members of a given population. The fifteen members of a population that were most "fit" as defined by $E(\Phi)=E_1(\Phi)+\gamma E_2(\Phi)$ randomly exchanged a subset of parameters ("mating"), generating 30 "offspring" for the new population. Five members of the new population were modified by random perturbations of randomly selected components in their parameter set ("mutation"). The initial population of 30 parameter sets was determined randomly, using a uniform distribution in parameter space, over the parameter range discussed below. Ideally one desires more than 30 members for a population, but computational limitations (in running the forward-propagation model) dictated the population size considered here. Many more details on GA implementation may be found in Ref. 29. In Table I we present a comparison between the parameters Φ extracted via the GA and estimates of channel parameters based on estimations made at the experiment site. We also present in Table I the search range considered by the GA for each of the channel parameters. We see from Table I that the water depth agrees with on-site measurements to within a 0.8% difference; the water sound speed and the bottom sound speed agree with on-site measurements to within 0.05% and 2.9%, respectively.

Although seven channel parameters were employed in the above study, it is anticipated that some parameters may be more important than others. It is of interest to reduce the number of parameters considered in the forward model, to accelerate computations such as those needed for the GA. We performed GA-based parameter inversion ten times, with ten distinct initial GA populations (determined randomly), using a uniform distribution over the range of channel parameters as indicated in Table I. In Fig. 5 we present the parameters extracted for each of these ten GA computations. The results indicate that the attenuation of the bottom is not a critical parameter since the multiple GA runs yield substantial variation over this parameter, while yielding comparable final time-reversal-image quality. These results also underscore the GA robustness to selection of the initial parameter-space population.

We note that there have been other wideband experiments performed with the goal of inverting channel parameters. For example, the Yellow Shark experiment[33–35] inverted for the properties of the sea bottom, assuming knowledge of the depth and sound-speed profile of the waveguide. It is also important to emphasize that these previous measurements were performed in transmission, as opposed to the reflection measurements considered here, and we have inverted for the channel depth and sound speed, as well as the bottom parameters.

C. Time-reversal imaging

Having determined the channel parameters, it is of interest to examine the data quality before and after time-reversal

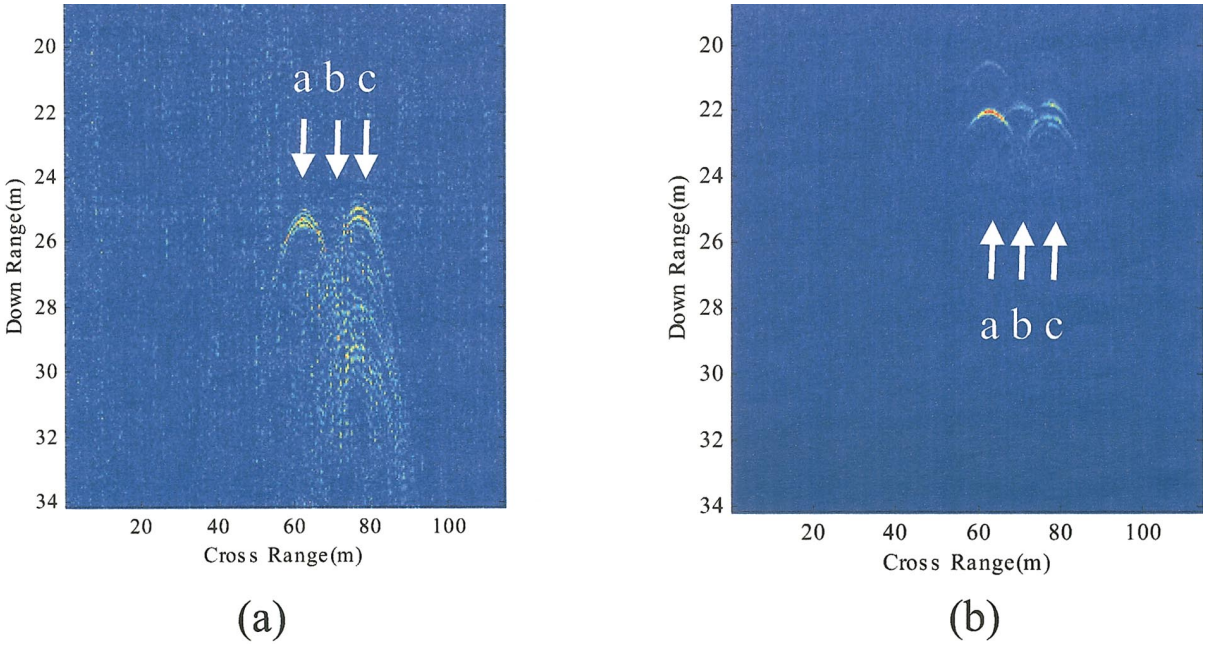


FIG. 6. Measured backscattered data for three targets at the bottom of a bay. The bay parameters are described in Table I. The vertical array measures the scattered response of the targets, at a fixed target-sensor orientation. Multiple target-sensor orientations are considered by moving the sensor horizontally along a line. The sensor path is parallel to the horizontal axis of these figures. In the figures the horizontal axis corresponds to the different positions of the sensor along the synthetic linear aperture, and the vertical axis represents range away from the sensor. The three targets are labeled ''a,'' ''b,'' and ''c.'' (a) original ''raw'' measured data, (b) data after time-reversal imaging.

imaging (using channel parameters determined as above by the GA). As discussed above, the acoustic system was used to collect data for multiple positions along a linear path (Fig. 4) constituting a synthetic linear horizontal aperture (and a physical vertical aperture). Time-reversal imaging is performed for each sensor position along the linear path. Subsequently, all of the data are used to synthesize a linear aperture, with which we obtain a synthetic-aperture-sonar (SAS) image[31] of the targets. In Fig. 6 we plot the data as a function of the position along the linear aperture, before SAS processing. In particular, the horizontal axis represents the position of the vertical array (along the linear synthetic aperture) and the vertical axis represents range from the sensor array (related to time delay in the observed data).

Figure 6(a) represents the ''raw'' measured data, averaged across all elements of the vertical array (with the time-domain signal converted to a function of range using the known sound speed, but not accounting for the channel). In Fig. 6(b) we present the time-reversal results observed at a depth of 8.2 m, corresponding to the nominal target depth. Specifically, in Fig. 6(b) we plot $\hat{I}_s(i,j;\Phi)$ as a function of j (range index) for fixed depth i (corresponding to 8.2 m) and channel parameters, Φ, determined via the GA inversion (Table I). The targets are labeled ''a,'' ''b,'' and ''c'' from left to right. The most important difference between the data plotted in Figs. 6(a) and (b) is that the time-reversal results yield the correct down range extent of the targets (1 to 2 meters) while the raw data yields targets that appear approximately 6 meters long. The distorted range of the targets in the raw data is mainly due to channel multipath effects and, to a lesser extent, the elastic nature of the target scattering. The hyperbolic shape of the highlights in Fig. 6 is caused by increasing and decreasing distance between the sensor and scattering centers as a function of position along the linear synthetic aperture (corresponding to cross-range position in Fig. 6).

In Fig. 6 each cross-range position (horizontal axis) corresponds to a single position of the vertical sensor array along the synthetic linear aperture. We may now take the data for all cross-range measurement positions and synthesize a linear aperture, yielding a synthetic-aperture sonar (SAS) image. We first consider time-reversal SAS. Let $s_n(t,z,r)$ represent the time-reversal-generated signal at time t, range r and depth z, associated with vertical array position n along the linear synthetic aperture (i.e., the nth vertical array position along the horizontal linear synthetic aperture). As indicated in Fig. 7, a scattering center at (x,y,z) corresponds to a distinct range r from each vertical-array position. The value of the time-reversal SAS image at position (x,y,z) is a linear superposition in n of the $s_n[t,z,r(x,y)]$, where we emphasize that the range associated with vertical array n is dependent on the cross-sectional position (x,y). To emphasize

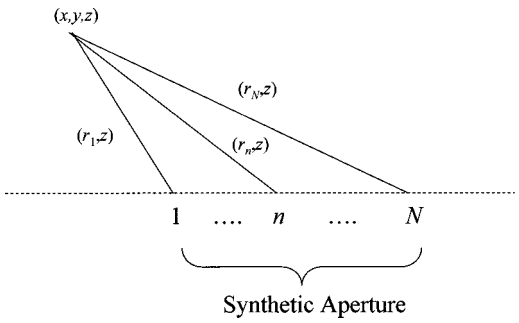


FIG. 7. Depiction of synthetic-aperture-sonar (SAS) image formation, from N different positions of the same vertical array, along a linear path.

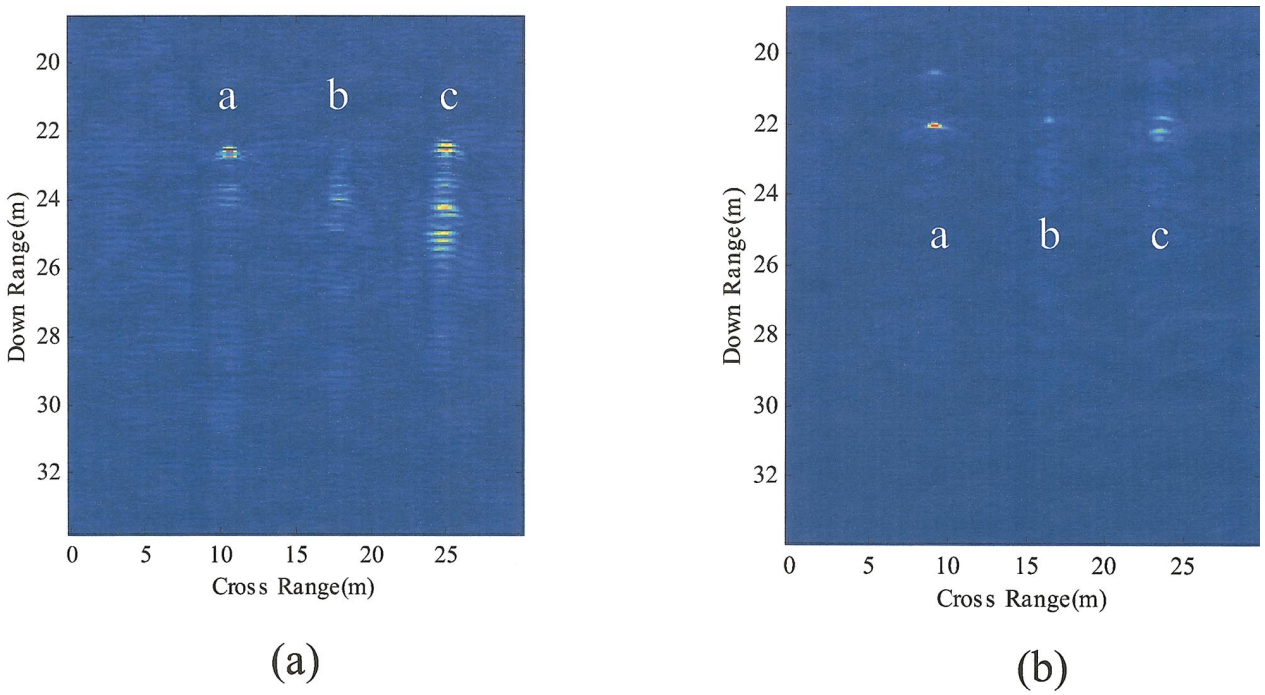


FIG. 8. Results of utilizing the data in Fig. 6 to constitute a synthetic aperture, from which we obtain these synthetic-aperture-sonar (SAS) images. (a) SAS image based on original data without accounting for effects of waveguide, (b) SAS image based on time-reversal imaging of each sensor position.

wavefront scattering at the target scattering centers, we consider the time-reversal image at $t=0$, for a fixed depth z (generating a two-dimensional image). In general one may generate a four-dimensional space–time SAS time-reversal image.

For comparison, we also show results of SAS imaging without time-reversal. In this case $s_n(z,r)$ corresponds to the depth-range-dependent image associated with vertical-array position n, where the reflections at the channel interfaces are ignored and the signal at (z,r) is determined by summing the contributions from all elements of the vertical array. Here the channel-induced multiple reflections are ignored and straight-ray propagation is assumed from each vertical-array element to the point being imaged. Note that the time delay has been used to generate $s_n(z,r)$; in particular time has been converted via the sound speed to distance from the nth vertical sensor. To produce a SAS image at (x,y,z) we sum in n the $s_n[z,r(x,y)]$ as in the time-reversal case, but now the subsequent image is time-independent.

In Fig. 8 we present SAS images of the data addressed in Fig. 6, with and without time-reversal imaging. In Fig. 8(a) the SAS imaging is performed without time reversal (i.e., not accounting for the channel), and in Fig. 8(b) the SAS image is generated using the time-reversal data from Fig. 6(b). The time-reversal SAS results yield significantly improved focusing of the target scattering centers. Both images correspond to a depth of 8.2 m, the nominal target position. The importance of accounting for the channel when imaging has also been discussed in Ref. 36.

Comparison of Figs. 8(a) and (b) reveals a key distinction between conventional SAS and time-reversal SAS, particularly in the context of target "c," which based on its elastic composition is highly resonant. The initial reflection from a target reveals its range, and subsequent resonant scattering between target scattering centers is observed at the sensor later in time. In conventional SAS time is mapped to range, using the speed of sound, and therefore the resonant characteristics of target "c" yield a strong resonant signature *behind* target "c" (i.e., further from the sensor than the target location). In time-reversal SAS, by contrast, we yield a four-dimensional image, and at $t=0$ we concentrate on the initial localized (nonresonant) wavefront scattering. Therefore, in Fig. 8(b) target "c" does not have an anomalous resonant "tail" behind it. This anomalous phenomenon in Fig. 8(a) makes target "c" appear to have a physical extent much larger than reality, while this anomaly is removed in the time-reversal image of Fig. 8(b).

V. CONCLUSIONS

The time-reversal method has been considered for imaging of an elastic target situated in an acoustic waveguide. Time-reversal imaging extracts the target scattering centers and with time-reversal imaging one recovers an approximation to the scattering-center time dependence. If a strong scatterer is present in the domain of interest, the quality of the associated time-reversal image may be used to search for the channel parameters. In the work reported here such a search has been performed via a genetic algorithm (GA), with example results presented based on measured data. The measured bay data considered here corresponded to a "real" but relatively benign scenario (e.g., approximately flat bottom). Future studies are needed to assess the effects of surface roughness on time-reversal-image quality and on the subsequent classification performance.

Having demonstrated time-reversal imaging, the next step is consideration of target classification based on such imagery. The quality of the time-reversal imagery is often

strongly dependent on the degree to which the Green's function used in the time-reversal imager matches the actual channel properties. Estimation of the channel Green's function has been performed here via a GA, but a subsequent classifier must take into account inaccuracies in channel-parameter estimation that will manifest distortion in the subsequent imagery. Another important area of future study involves feature selection from the time-reversal imagery, in a four-dimensional feature space.

ACKNOWLEDGMENTS

The authors are grateful to C. F. Gaumond, A. J. Romano, and E. G. Williams for valuable discussions concerning appropriate Helmholtz integral and linear operator representations. This research has been sponsored, in part, by the U.S. Office of Naval Research (Code 32CM and Code 31).

[1] R. P. Gorman and T. J. Sejnowski, ''Learned classification of sonar targets using a massively parallel network,'' IEEE Trans. Acoust., Speech, Signal Process. **36**, 1135–1140 (1988).
[2] B. A. Telfter, H. H. Szu, and G. Dobeck, ''Adaptive time-frequency classification of acoustic backscatter,'' Proc. SPIE Int. Symp. Aerospace/Defense Sensing Contr, Orlando, FL, April 1995.
[3] G. Goo and W. L. Au, ''Detection and identification of buried objects in shallow water,'' Proc. SPIE Int. Symp. Aerospace/Defense Sensing Contr, Orlando, FL, pp. 201–214, April 1996.
[4] P. H. Carter and G. Dobeck, ''Classification of acoustic backscatter using the generalized target description,'' Proc. SPIE Int. Symp. Aerospace/Defense Sensing Contr, Orlando, FL, pp. 190–200, April 1996.
[5] N. Intrator, Q. Q. Huynh, and G. Dobeck, ''Feature extraction from backscattered signals using wavelet dictionaries,'' Proc. SPIE Int. Symp. Aerospace/Defense Sensing Contr, Orlando, FL, pp. 183–190, April 1997.
[6] L. L. Burton and G. Dobeck, ''Active sonar target imaging and classification system,'' Proc. SPIE Int. Symp. Aerospace/Defense Sensing Contr, Orlando, FL, pp. 19–33, April 1997.
[7] M. R. Azimi-Sadjadi, Q. Huang, and G. Dobeck, ''Underwater target classification using multiaspect fusion and neural network,'' Proc. SPIE Int. Symp. Aerospace/Defense Sensing Contr, Orlando, FL, pp. 334–341, April 1998.
[8] M. R. Azimi-Sadjadi, D. Yao, Q. Huang, and G. Dobeck, ''Underwater target classification using wavelet packets and neural networks,'' IEEE Trans. Neural Netw. **11**, 784–794 (2000).
[9] H. W. Liu and L. Carin, ''Class-based target classification in shallow water channel based on Hidden Markov Model,'' Proc. IEEE International conference on acoustics, speech and signal processing, pp. 2889–2892, Orlando, FL, May 2002.
[10] H. Liu, P. Runkle, L. Carin, T. Yoder, T. Giddings, L. Couchman, and J. Bucaro, ''Classification of distant targets situated near channel bottoms,'' to appear in JASA, 2004.
[11] P. R. Runkle, P. K. Bharadwaj, L. Couchman, and L. Carin, ''Hidden Markov models for multi-aspect targets classification,'' IEEE Trans. Signal Process. **47**, 2035–2040 (1999).
[12] P. K. Bharadwj, P. R. Runkle, and L. Carin, ''Target identification with wave-based matched pursuits and Hidden Markov Models,'' IEEE Trans. Antennas Propag. **47**, 1543–1554 (1999).
[13] N. Dasgupta, P. K. Bharadwaj, L. Couchman, and L. Carin, ''Dual hidden Markov model for characterizing wavelet coefficients from multi-aspect scattering data,'' Signal Process. **81**, 1303–1316 (2001).
[14] M. K. Broadhead, ''Broadband source signature extraction from underwater acoustics data with sparse environment information,'' J. Acoust. Soc. Am. **97**, 1322–1325 (1995).
[15] M. K. Broadhead, L. A. Pflug, and R. L. Field, ''Use of high order statistics in source signature estimation,'' J. Acoust. Soc. Am. **107**, 2576–2585 (2000).
[16] A. Sarkissian, ''Extraction of a target scattering response from measurements made over long ranges in shallow water,'' J. Acoust. Soc. Am. **102**, 825–832 (1997).
[17] 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).
[18] J. S. Kim, H. C. Song, and W. A. Kuperman, ''Adaptive time-reversal mirror,'' J. Acoust. Soc. Am. **109**, 1817–1825 (2001).
[19] M. Fink, ''Time-reversal acoustics,'' Phys. Today **50**, 34–40 (1997).
[20] J.-L. Thomas, F. Wu, and M. Fink, ''Time-reversal focusing applied to lithotrpsy,'' Ultrason. Imaging **18**, 106–121 (1996).
[21] L. Borcea, G. Papanicolaou, C. Tsogka, and J. Berryman, ''Imaging and time reversal in random media,'' Inverse Probl. **18**, 1247–1279 (2002).
[22] P. Blomberg, G. Papanicolaou, and H. K. Zhao, ''Super-resolution in time-reversal acoustics,'' J. Acoust. Soc. Am. **111**, 230–248 (2002).
[23] J. F. Lingevitch, H. C. Song, and W. A. Kuperman, ''Time reversed reverberation focusing in a waveguide,'' J. Acoust. Soc. Am. **111**, 2609–2614 (2002).
[24] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP Press, New York, 1994).
[25] P. Gerstoft, ''Inversion of seismoacoustic data using genetic algorithms and a posteriori probability distributions,'' J. Acoust. Soc. Am. **95**, 770–782 (1994).
[26] P. Gerstoft, ''Ocean acoustic inversion with estimation of a posteriori probability distributions,'' J. Acoust. Soc. Am. **104**, 808–819 (1998).
[27] S. E. Dosso, M. L. Yeremy, J. M. Ozard, and N. R. Chapman, ''Estimation of ocean bottom properties by matched-field inversion of acoustic field data,'' IEEE J. Ocean. Eng. **18**, 232–239 (1993).
[28] A. M. Thode, G. L. D'Spain, and W. A. Kuperman, ''Matched-field processing, geoacoustic inversion, and source signature recovery of blue whale vocalization,'' J. Acoust. Soc. Am. **107**, 1286–1300 (2000).
[29] D. E. Goldberg, *Genetic Algorithms* (Addision-Wesley, Reading, MA, 1989).
[30] L. B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves* (IEEE Press, New York, 1996).
[31] B. Steinberg, *Microwave Imaging with Large Antenna Arrays* (Wiley, New York, 1983).
[32] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
[33] J. P. Hermand, ''Broad-band geoacoustic inversion in shallow water from waveguide impulse response measurements on a single hydrophone: Theory and experimental results,'' IEEE J. Ocean. Eng. **24**, 41–66 (1999).
[34] J. P. Hermand and P. Gerstoft, ''Inversion of broad-band multitone data from the YELLOW SHARK summer experiments,'' IEEE J. Ocean. Eng. **21**, 324–346 (1996).
[35] M. Siderius and J. P. Hermand, ''Yellow Shark Spring 1995: Inversion results from sparse broadband acoustic measurements over a highly range-dependent soft clay layer,'' J. Acoust. Soc. Am. **106**, 637–651 (1999).
[36] R. K. Brienzo and W. S. Hodgkiss, ''Broad-band matched-field processing,'' J. Acoust. Soc. Am. **94**, 2821–2831 (1993).

Inversion of guided-wave dispersion data with application to borehole acoustics[a)]

Henning Braunisch[b)]
Department of Electrical Engineering and Computer Science and Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139-4307

Tarek M. Habashy, Bikash K. Sinha, and Jahir Pabon
Schlumberger-Doll Research, Old Quarry Road, Ridgefield, Connecticut 06877-4108

Jin A. Kong
Department of Electrical Engineering and Computer Science and Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139-4307

(Received 17 November 2000; revised 16 May 2003; accepted 3 September 2003)

The problem of inferring unknown geometry and material parameters of a waveguide model from noisy samples of the associated modal dispersion curves is considered. In a significant reduction of the complexity of a common inversion methodology, the inner of two nested iterations is eliminated: The approach described does not employ explicit fitting of the data to computed dispersion curves. Instead, the unknown parameters are adjusted to minimize a cost function derived directly from the determinant of the boundary condition system matrix. This results in an efficient inversion scheme that, in the case of noise-free data, yields exact results. Multimode data can be simultaneously processed without extra complications. Furthermore, the inversion scheme can accommodate an arbitrary number of unknown parameters, provided that the data have sufficient sensitivity to these parameters. As an important application, we consider the sonic guidance condition for a fluid-filled borehole in an elastic, homogeneous, and isotropic rock formation for numerical forward and inverse dispersion analysis. We investigate numerically the parametric inversion with errors in the model parameters and the influence of bandwidth and noise, and examine the cases of multifrequency and multimode data, using simulated flexural and Stoneley dispersion data. [DOI: 10.1121/1.1625683]

PACS numbers: 43.60.Pt, 43.20.Mv [ANN] Pages: 269–279

I. INTRODUCTION

A common picture for the propagation in waveguides comprises elementary waves bouncing between the boundaries of the structure, interfering constructively only for certain directions of the elementary wave vectors.[1] This determines the set of possible waveguide modes and their propagation constants. The propagation characteristics in the waveguide are therefore related not only to the geometrical dimensions of the structure but also to the properties of the materials constituting the waveguide. Interpreting differently, the propagating field of an open waveguide penetrates into the surrounding medium, probing its constitutive parameters that influence the dispersive characteristics of the waveguide.

This motivates the acquisition of dispersion data and inferring parameters of interest with the help of an appropriate inversion scheme. An advantage of this basic idea is the source independence of the dispersion in waveguides that circumvents the need for modeling the transmitters and receivers of the envisioned apparatus. A disadvantage is that dispersion curves can be measured only indirectly, i.e., the inversion has to be preceded by a processing of the measured waveforms in order to extract the dispersion information. This by itself can be difficult and prone to the introduction of uncertainties beyond those of the measurement. However, the advantage of source independence might justify the additional preprocessing step.

After describing a methodology for the inversion of general modal dispersion data we take the case of borehole acoustics as an example. We consider a logging tool equipped with transmitters and an array of receivers for recording time-domain waveforms that can be processed in order to separate the various arrivals and obtain samples of the pertinent dispersion curves.[2,3] Previous inversion methods include curve fitting procedures that are designed to minimize the error between the measured dispersion samples and synthetic data obtained from a forward model that uses an iterative modal search routine.[4,5] The same forward model could be utilized to approximate the partial derivatives of the synthetic data with respect to the model parameters that are required in a systematic inversion algorithm. Alternatively, these sensitivity coefficients can be computed by numerical quadrature of appropriate perturbation integrals.[4–7] Perturbation theory and a comparison of measured and computed dispersion curves plays an important role also in the interpretation of dispersion data influenced by nonlinear effects.[8–10] Kimball describes a processing scheme that is not

[a)]Portions of this work were presented in "Inversion of borehole dispersions for formation elastic moduli," *Proceedings of the IEEE International Ultrasonics Symposium*, San Juan, Puerto Rico, 22–25 October 2000, Vol. 1, pp. 551–556.

[b)]Electronic mail: henning.braunisch@intel.com

based on direct curve fitting but, again, computed dispersions are required explicitly.[11,12]

In Sec. II of the present work we suggest minimizing a cost function that is derived directly from the determinant representing the boundary conditions at the interfaces of a linear waveguide model.[13] This cost function can be evaluated rapidly, without the need for any iterations, and allows the possibility of processing multifrequency and multimode data in a potentially exact inversion. An arbitrary number of unknown parameters can be accommodated, provided that the data have sufficient sensitivity to these parameters. The theoretical investigations and numerical simulations carried out and described in this paper do not incorporate any influence of the logging tool and only the simplest case of a fluid-filled borehole in a homogeneous and isotropic elastic rock formation is examined. In Sec. III we give a suitable formulation of the relevant guidance condition. Then, in Sec. IV, we investigate the inversion with model parameter errors, the influence of bandwidth and noise, and the utilization of multifrequency and multimode data.

II. PARAMETRIC INVERSION OF GUIDED-WAVE MODAL DISPERSIONS

A. Modal dispersion curves

The guidance condition or characteristic equation for a two-dimensional waveguide structure invariant in the z direction and described by a parameter vector $\bar{x}$ containing geometrical and material constants can be written as

$$D(k_z,\omega,\bar{x})=0. \tag{1}$$

Here D is the determinant of the system matrix $\bar{\bar{L}}$ of the homogeneous linear system of equations that follows from matching the appropriate boundary conditions, k_z is the wave number in the direction of propagation, and ω is the angular frequency considered.

If we choose a smooth curve Ω in the ω domain (typically, but not necessarily, the positive real axis), then the roots of Eq. (1) for some $\omega=\omega_0\in\Omega$ constitute a set of modes. Choosing a particular mode by picking one of the roots, the dispersion relation $k_z(\omega,\bar{x})$ for this mode (with respect to Ω) is obtained by tracing the locus of the root in the k_z domain as ω moves away from ω_0 and along Ω.

This notion of dispersion leads directly to a numerical method for computing modal dispersion curves practically. A mode tracking procedure can be implemented with the help of, e.g., the iterative complex Newton–Raphson method.[14]

B. Inversion of modal dispersion data

We now turn to the inverse problem, i.e., an estimation of the N unknown elements of $\bar{x}$ from bandlimited, possibly noisy, samples of one or more dispersion curves. The number of parameters N to be determined can be less than the dimension of $\bar{x}$, e.g., in case some of the elements of $\bar{x}$ were obtained from other measurements.

Given M measured pairs (ω_i,k_{zi}) that satisfy

$$k_{zi}=k_z(\omega_i,\bar{x})+n_i,\quad i=1,2,...,M, \tag{2}$$

where the n_i represent the noise in the data and, as in the case of multimode data, the $k_z(\omega_i,\bar{x})$ may belong to different modes, one possible formulation of the problem aims at minimizing the cost function

$$\|\bar{e}_0(\bar{x})\|^2=\sum_{i=1}^{M}|k_z(\omega_i,\bar{x})-k_{zi}|^2. \tag{3}$$

The problem with this approach is that every single evaluation of Eq. (3) for varying $\bar{x}$, in whatever optimization method employed, requires the M roots $k_z(\omega_i,\bar{x})$, which can be determined exactly only by iteration (Sec. II A). If one tries to avoid these iterations by precomputing a look-up table of k_z for all possible $\bar{x}$, ω, and modes—which itself might be an extensive computational task and may require a prohibitively large database—then the necessary interpolations during the inversion would preclude an exact answer, even in the case of noise-free data. Furthermore, the implicit mode identification problem, i.e., relating each k_{zi} to the correct dispersion curve, may complicate the situation.

The above difficulties can be avoided by solving the inverse problem without resorting to the dispersion curves $k_z(\omega,\bar{x})$ explicitly. We therefore suggest pursuing the minimization of the "guidance mismatch," a measure of the error in satisfying the guidance condition Eq. (1), defined by

$$\|\bar{e}(\bar{x})\|^2=\sum_{i=1}^{M}|D(k_{zi},\omega_i,\bar{x})|^2. \tag{4}$$

It is obvious that the cost function given by Eq. (4), for the case of noise-free data, k_{zi}, can be made zero by varying $\bar{x}$, similar to Eq. (3). For noisy data, the least-squares problem can be solved by applying the Gauss–Newton method.[15] The partial derivatives in the Jacobian are typically replaced by finite differences unless the structure considered is simple enough so that the differentiations can be carried out analytically. While the cost function in Eq. (3) is of the same form as in curve fitting problems, in Eq. (4) the data k_{zi} and thus the noise influence $\bar{e}$ in a nonlinear fashion.

III. SONIC GUIDANCE CONDITION FOR A FLUID-FILLED BOREHOLE

A. Formulation for forward and inverse dispersion analysis

For linearized acoustic wave propagation along a fluid-filled borehole in an elastic, homogeneous, and isotropic rock formation (Fig. 1),

$$\bar{x}=[a,v_s,v_p,v_f,\varrho_{fr}], \tag{5}$$

with a the borehole radius, v_s the shear wave velocity in the formation, v_p and v_f the compressional wave velocities in the formation and the fluid, respectively, and $\varrho_{fr}=\varrho_f/\varrho$ the ratio of mass densities of the fluid and the formation (relative fluid mass density, usually less than one).

The four scalar boundary conditions at an interface between two elastic media where transverse slip may occur require continuity of the normal component of the particle velocity vector $\bar{v}$, and continuity of the normal component and vanishing of the tangential components of the traction vector $\bar{\bar{\tau}}\cdot\hat{n}$, where $\bar{\bar{\tau}}$ is the symmetric stress tensor and $\hat{n}$ the

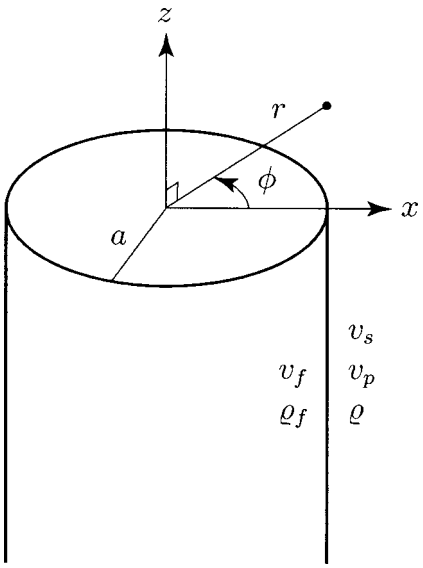


FIG. 1. A fluid-filled borehole in an elastic, homogeneous, and isotropic rock formation, forming a circularly cylindrical fluid–solid interface.

unit vector normal to the interface.[16] An inviscid fluid can be considered as a special case of an elastic medium with zero shear wave velocity. In this case, the stress tensor specializes to

$$\bar{\bar{\tau}} = -p\bar{\bar{I}}, \tag{6}$$

with p the fluid pressure and $\bar{\bar{I}}$ the identity tensor. Thus, the boundary conditions at the fluid–solid interface in Fig. 1 are continuity of $\hat{r}\cdot\bar{v}$ and

$$\hat{r}\cdot\bar{\bar{\tau}}\cdot\hat{r} = -p, \tag{7}$$

$$\hat{\phi}\cdot\bar{\bar{\tau}}\cdot\hat{r} = 0, \tag{8}$$

$$\hat{z}\cdot\bar{\bar{\tau}}\cdot\hat{r} = 0. \tag{9}$$

Continuing directly with the formulation in Ref. 17, we adopt an $e^{ik_z z - i\omega t}$ space–time dependence by using the identity relating the Hankel functions of the first and second kind,[18]

$$H_n^{(1)}(\zeta^*) = [H_n^{(2)}(\zeta)]^*, \tag{10}$$

where ζ is a generic complex variable, the asterisk denotes complex conjugation, and $n=0,1,...$ is the azimuthal mode order. Some rearranging leads to the 4×4 boundary condition system matrix $\bar{\bar{L}}=[L_{ij}]$ with

$$L_{11} = k_{fr}aJ_{n+1}(k_{fr}a) - nJ_n(k_{fr}a), \tag{11}$$

$$L_{21} = \varrho_{fr}(k_s a)^2 J_n(k_{fr}a), \tag{12}$$

$$L_{31} = 0, \tag{13}$$

$$L_{41} = 0; \tag{14}$$

$$L_{12} = n - A_n(k_{pr}a), \tag{15}$$

$$L_{22} = 2[n(n-1) + (k_z a)^2 + A_n(k_{pr}a)] - (k_s a)^2, \tag{16}$$

$$L_{32} = 2n[1 - n + A_n(k_{pr}a)], \tag{17}$$

$$L_{42} = 2ik_z a[A_n(k_{pr}a) - n]; \tag{18}$$

$$L_{13} = -ik_z a, \tag{19}$$

$$L_{23} = 2ik_z a[n+1-B_n(k_{sr}a)], \tag{20}$$

$$L_{33} = ik_z a[2(n+1) - B_n(k_{sr}a)], \tag{21}$$

$$L_{43} = (k_s a)^2 - 2(k_z a)^2 - nB_n(k_{sr}a); \tag{22}$$

$$L_{14} = n, \tag{23}$$

$$L_{24} = 2n[n-1-A_n(k_{sr}a)], \tag{24}$$

$$L_{34} = (k_{sr}a)^2 - 2n(n-1) - 2A_n(k_{sr}a), \tag{25}$$

$$L_{44} = -nik_z a; \tag{26}$$

where $J_n(\zeta)$ denotes the Bessel function of the first kind and order n and we introduced the auxiliary functions

$$A_n(\zeta) = \zeta H_{n+1}^{(1)}(\zeta)/H_n^{(1)}(\zeta), \tag{27}$$

$$B_n(\zeta) = \zeta H_n^{(1)}(\zeta)/H_{n+1}^{(1)}(\zeta) = \zeta^2/A_n(\zeta), \tag{28}$$

in order to avoid infinite values as $k_{\alpha r}\to 0$, $\alpha = s,p$. The normalized radial wave numbers $k_{\alpha r}$ satisfy the dispersion relations

$$k_{\alpha r}^2 = k_\alpha^2 - k_z^2, \quad \alpha = s,p,f, \tag{29}$$

where $k_\alpha = \omega/v_\alpha$. For proper modes, i.e., modes that are bound to the surface of the open waveguide (the fluid–formation interface) and that are either radially evanescent or represent an outgoing, radially decaying wave and that thus satisfy the radiation condition at infinity, we have $\mathrm{Re}\{k_{\alpha r}\} \geqslant 0$ and $\mathrm{Im}\{k_{\alpha r}\} > 0$, $\alpha = s,p$. The choice of sign for k_{fr} only contributes a factor of $(-1)^n$ to the determinant and thus does not influence the location of the roots of the characteristic equation.

Since we are interested in minimizing the cost function in Eq. (4), we should remove zeros of the determinant of $\bar{\bar{L}}$ that do not correspond to guided-wave modes in order to avoid the introduction of unwanted minima.

For $n>0$, column 1 of $\bar{\bar{L}}$ as given by Eqs. (11)–(14) vanishes as $k_{fr}\to 0$. Even if, in an inversion, the component v_f of $\bar{x}$ is not varied because this parameter might already be known (for a water-filled borehole, $v_f = 1500$ m/s), this can be a problem when computing dispersion curves. The zero can be removed by dividing column 1 by $k_{fr}a$.

For all n and as $k_{sr}\to 0$, the determinant based on the above formulation vanishes in a more subtle way. In view of

$$\lim_{\zeta\to 0} A_n(\zeta) = 2n, \tag{30}$$

$$\lim_{\zeta\to 0} B_n(\zeta) = 0, \tag{31}$$

we find that column 3 and column 4 become linearly dependent as $k_s \to k_z$. By adding

$$\frac{nL_{i3}}{ik_z a}$$

to L_{i4} and dividing the result by its third element the following new formulation for the fourth column of $\bar{\bar{L}}$ is obtained, as a replacement of Eqs. (23)–(26):

$$L_{14} = 0, \tag{32}$$

$$L_{24}=n\left[1-\frac{(k_{sr}a)^2-(n-2)B_n(k_{sr}a)}{4n+(k_{sr}a)^2-2A_n(k_{sr}a)-nB_n(k_{sr}a)}\right], \tag{33}$$

$$L_{34}=1, \tag{34}$$

$$L_{44}=\frac{ni}{k_z a}\,\frac{nB_n(k_{sr}a)-(k_{sr}a)^2}{4n+(k_{sr}a)^2-2A_n(k_{sr}a)-nB_n(k_{sr}a)}. \tag{35}$$

The fourth-order determinant with $L_{14}=L_{31}=L_{41}=0$ and $L_{34}=1$ can then be evaluated by expanding down the first column of the matrix and across the first row of the two 3×3 minors, yielding

$$\begin{aligned}D=L_{11}\{&L_{22}[(L_{33}L_{44})-L_{43}]+L_{23}[L_{42}-(L_{32}L_{44})]\\&+L_{24}(L_{32}L_{43}-L_{33}L_{42})\}-L_{21}\{L_{12}[(L_{33}L_{44})-L_{43}]\\&+L_{13}[L_{42}-(L_{32}L_{44})]\}.\end{aligned} \tag{36}$$

B. High-frequency limit

We can verify our formulation by deriving the equation for the Scholte wave velocity v_{Sch} (the speed of a surface wave along a *planar* fluid–solid interface), where

$$k_z=k_{\mathrm{Sch}}=\frac{\omega}{v_{\mathrm{Sch}}} \tag{37}$$

is expected to satisfy the guidance condition for the fluid-filled bore hole in the high-frequency limit, $\omega\to\infty$, when all wave numbers become so large that the local curvature of the fluid–solid interface is negligible. Using

$$A_n(\zeta)\sim-i\zeta, \tag{38}$$

$$B_n(\zeta)\sim i\zeta, \tag{39}$$

as $|\zeta|\to\infty$ and

$$\frac{J_{n+1}(k_{fr}a)}{J_n(k_{fr}a)}\sim\pm i, \tag{40}$$

as $\mathrm{Im}\,k_{fr}\to\pm\infty$, we obtain, as $\omega\to\infty$ and for $\mathrm{Im}\,k_{fr}>0$,

$$(k_z^2-k_{sr}^2)^2+4k_{sr}k_{pr}k_z^2+\varrho_{fr}k_s^4k_{pr}/k_{fr}=0, \tag{41}$$

which, as expected, is independent of n, a, and ω (the frequency ω can be dropped from the equation), and that checks with Ref. 19. It is noted that, for real v_{Sch} and v_f, Eq. (40) implies $v_{\mathrm{Sch}}<v_f$, and—more stringent—for the surface wave to be bound to the fluid–solid interface,

$$v_{\mathrm{Sch}}<v_{s,p,f}, \tag{42}$$

is required.

IV. PARAMETRIC INVERSION OF BOREHOLE DISPERSIONS

In the following we apply the inversion method introduced in Sec. II B to the physical model described by the determinant in Sec. III A. Although this example is drawn from acoustics, the methodology applies equally well to any other kind of waveguide; different waveguide structures can be examined by considering the appropriate guidance conditions. In the rest of the paper $\omega=2\pi f$ is real and positive.

TABLE I. Parameters for a water-filled borehole in a fast and a slow formation. Given also are the resulting Scholte and tube wave velocities v_{Sch} and v_t, respectively, and the Stoneley cutoff frequencies $f_{c,\mathrm{St}}$.

	a (cm)	v_s (m/s)	v_p (m/s)	v_f (m/s)	ϱ/ϱ_f	v_{Sch} (m/s)	v_t (m/s)	$f_{c,\mathrm{St}}$ (kHz)
Fast	10.16	2032	3658	1500	2.350	1439.6	1351.5	
Slow	10.16	508	1890	1500	2.054	452.2	655.0	0.835

We assume practically relevant parameters for a fast and a slow formation considered previously;[17] they are repeated in Table I for reference. A formation is classified as fast or slow depending on whether the shear wave velocity in the formation, v_s, is greater or less than the fluid wave velocity v_f. The Scholte wave velocities v_{Sch} obtained iteratively from Eq. (41) are also given in Table I and found to be in compliance with Eq. (42).

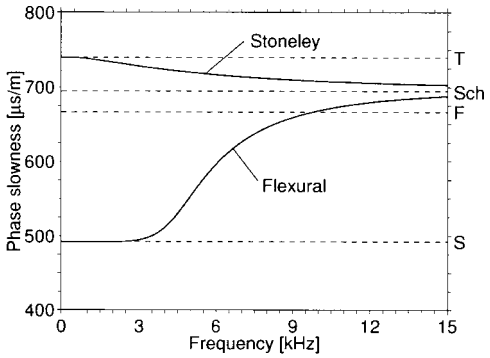
A. Inversion with errors in the model parameters

We now concentrate on the inversion for the formation shear wave velocity v_s, a quantity of particular practical interest in the exploration of hydrocarbons, and investigate the effect of systematic errors introduced into the model parameters borehole radius a and formation compressional wave velocity v_p, where in the first case the exact v_p is assumed to be known and in the second case the exact a. The fluid compressional wave velocity v_f and relative fluid mass density ϱ_{fr} are used without error.

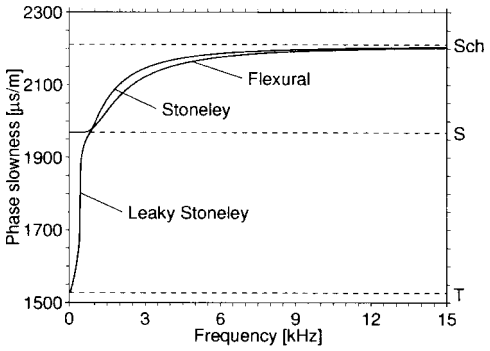
1. Flexural dispersion

For noise-free data, it is sufficient to provide a single "measurement" in order to determine the one parameter v_s ($M=N=1$). This point is taken from the dispersion of the lowest-order mode with $n=1$, known as the flexural mode, which can be excited by a dipole source placed on the borehole axis.[5,20] For the parameters in Table I, the flexural dispersions are plotted in Fig. 2 (similar curves can be found in Ref. 17), where phase slowness is defined as $\mathrm{Re}\{k_z\}/\omega$. The dashed horizontal lines labeled as "F," "S," and "Sch" show the slownesses corresponding to v_f, v_s, and v_{Sch}, respectively. "S" is the low-frequency asymptote of the flexural dispersion while "Sch" is approached as $\omega\to\infty$.

The results in Fig. 3 for four different frequencies at which the sample of the dispersion curves in Fig. 2 was obtained show that if in the inversion the assumed borehole radius is larger than the true borehole radius (leading to the dispersions in Fig. 2), then the inverted shear wave velocity exhibits a positive error; the opposite is true for a too small borehole radius. Starting off with a relatively small absolute value at the lowest frequencies considered (4 kHz for the fast formation and 1 kHz for the slow formation), the error in inverted v_s initially grows as the frequency increases and then tends to smaller values again. This is consistent with the asymptotic behavior of the flexural dispersion curve, which for both low and high frequencies approaches values that are independent of a. The relative errors in inverted shear wave velocity for the slow formation are found to be about ten times smaller than for the fast formation, where in the latter

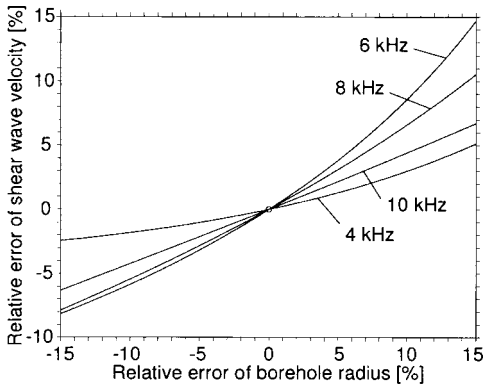


(a) Fast formation.

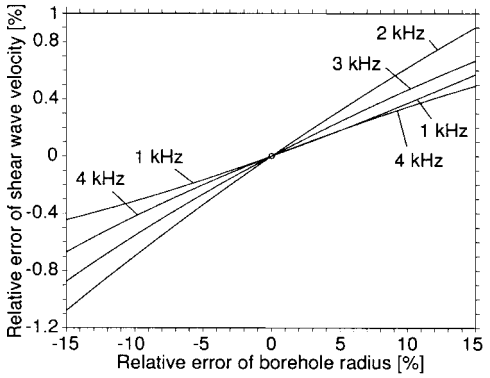


(b) Slow formation.

FIG. 2. Flexural and Stoneley dispersions for a water-filled borehole.

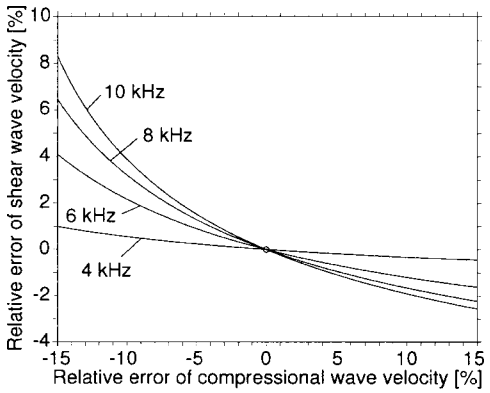


(a) Fast formation.

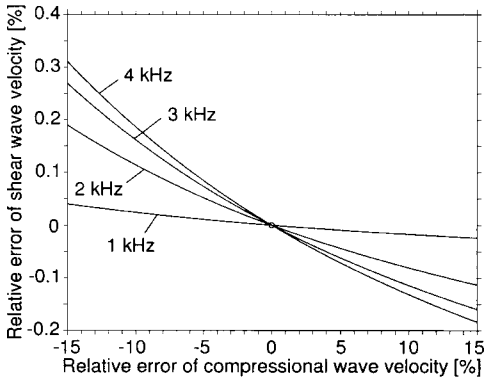


(b) Slow formation.

FIG. 3. Relative error in inverted formation shear wave velocity for varying relative error of an assumed borehole radius (flexural dispersion).



(a) Fast formation.



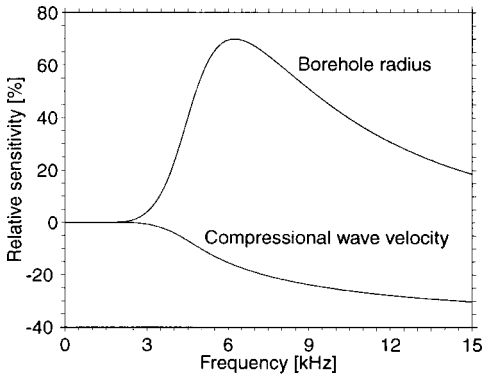
(b) Slow formation.

FIG. 4. Relative error in inverted formation shear wave velocity for varying relative error of assumed formation compressional wave velocity (flexural dispersion).

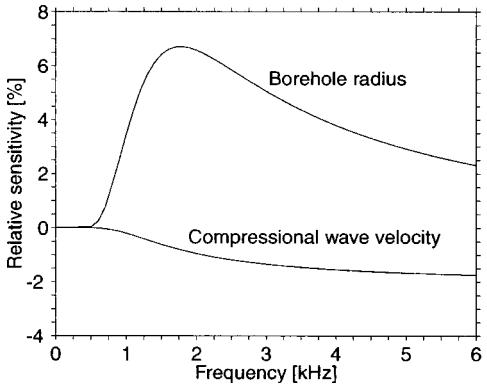
case the ratio of error in shear wave velocity to error in borehole radius can be close to one. This shows that in the slow formation an accurate estimate of v_s can be obtained, even when the value of a is uncertain. In the fast formation a good estimate of the borehole radius is required to obtain an accurate inversion result. On the other hand, reversing the roles of v_s and a, if v_s in the fast formation is known with little error then the dispersion information for intermediate frequencies could be used to find an accurate estimate of a. This would not work for the slow formation, where the flexural dispersion is much less sensitive to the borehole radius.

Similar results are found for the inversion of v_s when the assumed formation compressional wave velocity v_p exhibits an error (Fig. 4). Positive errors in v_p yield a result for v_s that is smaller than the exact value. The magnitude of the error of inverted v_s is found to be increasing with frequency. This is consistent with the fact that the low-frequency asymptote of the flexural dispersion is independent of v_p while the Scholte wave velocity v_{Sch} depends on v_p. In fact, given v_s, v_f, ϱ_{fr}, and v_{Sch}, Eq. (41) can be rearranged to give an explicit expression for v_p. The result is

$$v_p^{-2} = v_{\mathrm{Sch}}^{-2} - \frac{(2v_{\mathrm{Sch}}^{-2} - v_s^{-2})^4}{[4v_{\mathrm{Sch}}^{-2}(v_{\mathrm{Sch}}^{-2} - v_s^{-2})^{1/2} - \varrho_{fr} v_s^{-4}(v_{\mathrm{Sch}}^{-2} - v_f^{-2})^{-1/2}]^2}. \tag{43}$$



(a) Fast formation.



(b) Slow formation.

FIG. 5. Frequency-dependent relative sensitivity of inverted formation shear wave velocity to model error in borehole radius or formation compressional wave velocity (flexural dispersion).

As with the dependence of inverted shear wave velocity on the error in the borehole radius, errors in formation compressional wave velocity have a much larger influence on the inverted v_s in the case of the fast formation. Comparing Fig. 4 with Fig. 3, one notes that, in general, the sensitivity of the flexural dispersion data to v_p is smaller than the sensitivity to a. Thus the flexural dispersion for the slow formation is found to be insensitive to both a and v_p while in the case of the fast formation it is sensitive to a and moderately sensitive to v_p. The shear wave velocity v_s acting as the low-frequency asymptote has, of course, a strong influence on both dispersions. Similar conclusions can be found in Ref. 5, where a perturbation model is used to calculate the sensitivity of the flexural dispersion curve to small changes of the parameters.

While from Fig. 3 and Fig. 4 we can obtain detailed information about the error of the inversion as it depends nonlinearly on errors of the underlying model, with frequency as a parameter, in Fig. 5 we show results for continuously varying frequency. The relative sensitivities plotted are defined, for arbitrary frequency, as the slopes of the corresponding curves in Fig. 3 and Fig. 4 at the origin that represents the case of an inversion without error in the model. These relative sensitivities thus give the error of the otherwise exact inversion of shear wave velocity for small errors in the model parameters. For example, the relative sensitivity with respect to the borehole radius in the fast formation [Fig. 5(a)] peaks near 6 kHz at around 70%. This means that for an inversion with 1% error in the borehole radius we would expect a result with 0.7% error in shear wave velocity, assuming that all other parameters are known exactly.

2. Stoneley dispersion

The Stoneley mode[21] is the lowest-order axisymmetric mode ($n=0$) and can be excited by a monopole source placed on the borehole axis. The dispersion curves for the fast and the slow formation are plotted in Fig. 2 (see also Ref. 17). At the high-frequency end of the spectrum, the Stoneley mode approaches the same asymptotic value as the flexural dispersion, given by the Scholte wave velocity v_{Sch} [recall that Eq. (41) is independent of the azimuthal mode order n]. The dashed horizontal line labeled "T" in each subfigure of Fig. 2 indicates the slowness corresponding to the tube wave velocity v_t. The tube wave velocity is the low-frequency asymptote of the Stoneley dispersion, and is given by[21,22]

$$v_t^{-2}=v_f^{-2}+\varrho_{fr}v_s^{-2}. \tag{44}$$

The resulting tube wave velocities for the parameters in Table I are listed in the same table for reference. According to Eq. (44), the tube wave velocity is always slower than the fluid compressional wave velocity v_f and, in a fast formation, slower than the formation shear wave velocity v_s [see the introduction to Sec. IV and Fig. 2(a)]. However, in a slow formation, $v_t>v_s$ is possible, namely if

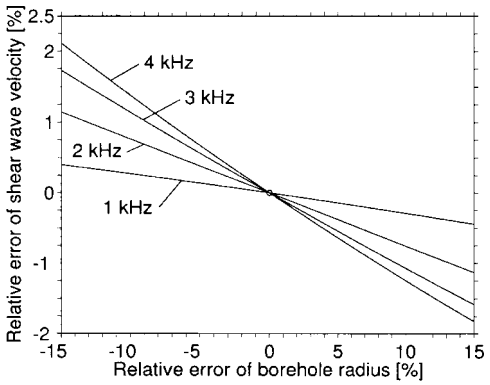
$$\frac{v_s}{v_f}<\sqrt{1-\varrho_{fr}}, \tag{45}$$

as follows from the definition Eq. (44) under the assumption $\varrho_{fr}<1$. Since for the high-frequency asymptote, $v_{\mathrm{Sch}}<v_s$ as stated in Eq. (42), under the condition Eq. (45) the Stoneley dispersion crosses the shear wave velocity at some cutoff frequency below which the wave is leaky, with k_{sr} and k_z in Eq. (29) complex. This phenomenon can be observed in Fig. 2(b), with the cutoff frequency given in Table I.

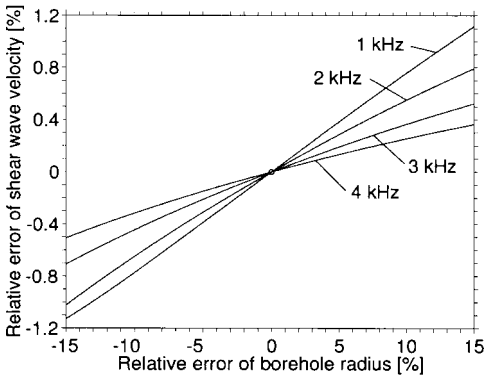
The results in Fig. 2 show that the qualitative behavior of the Stoneley dispersion in the two formations considered is completely different. In the slow formation, the nonleaky portion of the Stoneley dispersion is close to the flexural dispersion, while this is not the case in the fast formation. While the phase slowness of the Stoneley mode in Fig. 2(a) is monotonically decreasing, it is monotonically increasing in Fig. 2(b).

The similarities and differences between Stoneley and flexural dispersions carry over to the results in Figs. 6 and 7 that are Stoneley results analogous to the results displayed in Figs. 3 and 4 for the flexural dispersion. Note the opposite signs in Fig. 6(a) as compared with Figs. 3 and 6(b). However, positive errors in formation compressional wave velocity lead to negative errors in inverted formation shear wave velocity in all cases (Figs. 4 and 7).

The frequency-dependent relative sensitivities for inverting Stoneley dispersion data that are depicted in Figs. 8 and 9 have a definition and meaning similar to the sensitivities plotted previously for the flexural dispersion (Fig. 5). The data of the four curves in Fig. 5 are shown again in Figs. 8 and 9 for comparison. Since the tube wave velocity v_t as

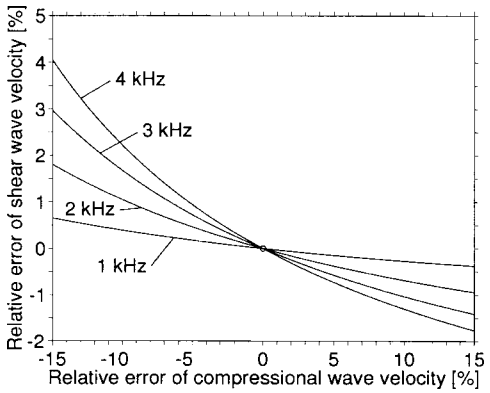


(a) Fast formation.

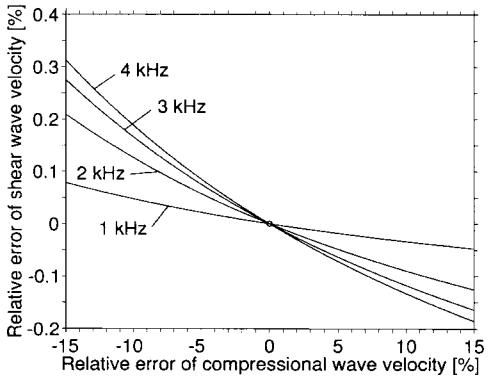


(b) Slow formation.

FIG. 6. Relative error in inverted formation shear wave velocity for varying relative error of an assumed borehole radius (Stoneley dispersion).



(a) Fast formation.



(b) Slow formation.

FIG. 7. Relative error in inverted formation shear wave velocity for varying relative error of assumed formation compressional wave velocity (Stoneley dispersion).

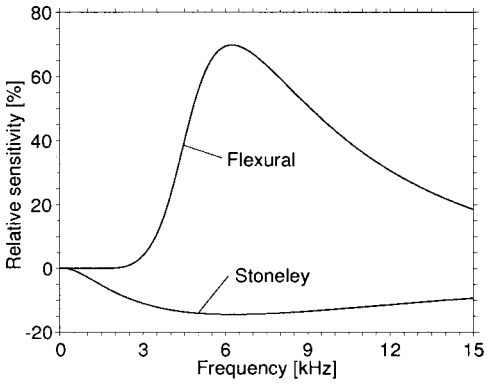
given by Eq. (44) is independent of borehole radius a and formation compressional wave velocity v_p, all Stoneley sensitivity curves shown go to zero as the frequency tends to zero. The sensitivity of inverted shear wave velocity to an error in the assumed borehole radius for the fast formation [Fig. 8(a)] is generally smaller in magnitude when inverting Stoneley dispersion data than when using flexural dispersion data, unless low frequencies are considered. The magnitude of the sensitivity of inverted shear wave velocity to an error in assumed v_p for the fast formation [Fig. 8(b)] is somewhat larger for the Stoneley dispersion as compared with the flexural dispersion. Setting the leaky Stoneley results aside, the most striking feature of Fig. 9 for the slow formation is the general closeness of the sensitivity curves for the Stoneley and the flexural dispersions. The sharp peaks of the relative sensitivities for the leaky Stoneley mode in Fig. 9 occur at that frequency where the dispersion curve in Fig. 2 goes through a very steep section.

B. Bandwidth and noise

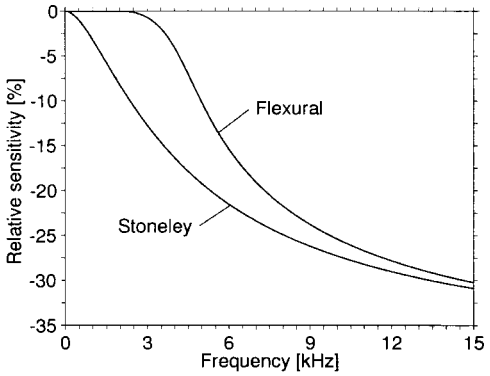
It is important to realize that sensitivity of the dispersion data to several parameters (e.g., v_s, v_p, and a for the fast formation) does not necessarily allow the simultaneous inversion of the data for these parameters. For example, if the data depend only on the ratio v_s/v_p and are sensitive to this ratio, then they will be sensitive to changes in v_s with v_p fixed, and *vice versa*, but a determination of both parameters will be impossible due to the nonuniqueness of the problem. Simultaneous inversion for N parameters requires, beside $M \geqslant N$, where M is the number of data points, a sufficient degree of independent information about the individual parameters in the data. It is reasonable to try adding more information to the data by sampling the dispersion curve with a larger bandwidth. We examine here the example of the fast formation and inverting v_s and v_p from flexural dispersion data.

We choose a fixed spacing $\Delta f = 200$ Hz between the samples that are centered about $f_c = 6.5$ kHz [a point to the right of the turning point of the dispersion curve in Fig. 2(a)], and, starting with $M = 2$, add points at both ends of the sampled interval, extending the bandwidth by 400 Hz at each step. For a fixed model error in formation compressional wave velocity v_p of -10% and with no noise added to the data, Table II shows the error in inverted shear wave velocity v_s with increasing bandwidth or, equivalently, the number of samples M. It is seen that the inversion result is practically independent of bandwidth and the resulting error, which could as well be read off from Fig. 4(a), is determined by the error in v_p. Table II also provides the number of iterations required in the application of the Gauss–Newton method. The initial guess for v_s is obtained directly from the data by taking it to be a few percent larger than the largest "measured" phase velocity.

In order to achieve a more accurate result for v_s and remove the uncertainty due to v_p, one can try to invert the

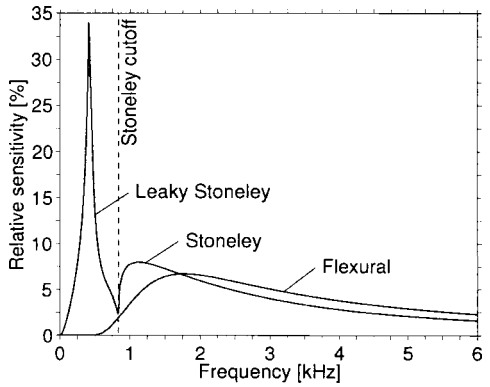


(a) Borehole radius.

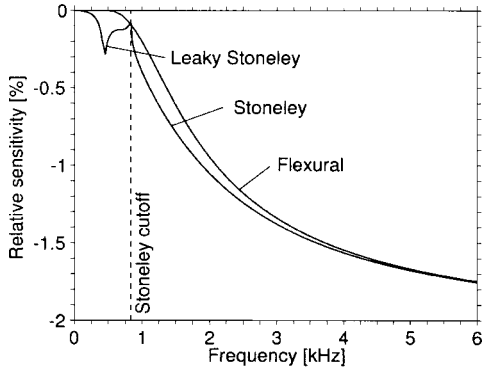


(b) Compressional wave velocity.

FIG. 8. Frequency-dependent relative sensitivity of inverted formation shear wave velocity to model error in borehole radius or formation compressional wave velocity using Stoneley or flexural dispersion data (fast formation).



(a) Borehole radius.



(b) Compressional wave velocity.

FIG. 9. Frequency-dependent relative sensitivity of inverted formation shear wave velocity to model error in borehole radius or formation compressional wave velocity using Stoneley or flexural dispersion data (slow formation).

data for the two parameters simultaneously ($N=2$). In the present noise-free case and for all bandwidths in Table II, this gives back the exact values for v_s and v_p listed in Table I and originally used to generate the dispersion curve in Fig. 2(a). This potentially exact inversion of dispersion data is an important feature of the methodology employed in this paper. However, the condition numbers of the inversion matrix in Table III show that the problem at hand is not well posed, i.e., the successful simultaneous inversion for v_s and v_p relies on the fact that the supplied data contain no errors. The condition number in the context here is defined as the ratio of the largest to the smallest of the eigenvalues of $\bar{\bar{J}}_0^T \cdot \bar{\bar{J}}_0$, where $\bar{\bar{J}}_0$ is the $2M \times N$ real Jacobian matrix at the start of the Gauss–Newton iteration. The reciprocal of this condition number is an indicator of how well defined the solution to the minimization problem is in terms of uniqueness. As seen from the data in Table III, increasing the bandwidth when sampling the flexural dispersion yields a considerable improvement of the condition number for the simultaneous inversion of v_s and v_p. But even for a 3 kHz bandwidth, after a decrease by about three orders of magnitude, the condition number is still much larger than the ideal unity.

The detrimental effect of ill conditioning in the presence of noise is demonstrated by the results in Table IV. Pseudo-random white Gaussian noise is added to the real-valued slowness data with a 3 kHz bandwidth, with the noise level defined as the relative root mean squared (rms) error. It is seen that only for the cases of 0% and 0.5% noise, reasonable inversion results are obtained. For a further increase of the noise level a rapid degradation is observed.

Returning to the case of inversion for formation shear wave velocity with fixed −10% error in formation compressional wave velocity (as in Table II), the results documented in Table V show that the inversion with $N=1$ is robust against noise in the data. Because only one particular realization of noise is added in each case, after an initial increase of the resulting error in inverted shear wave velocity with increasing noise level, we observe an improvement of the error; this would not be expected on average.

Similar results as described above are found when inverting the flexural dispersion data for the shear wave velocity and the borehole radius simultaneously (data not tabu-

TABLE II. Effect of increasing bandwidth on the inversion for formation shear wave velocity with −10% error in formation compressional wave velocity (fast formation, flexural dispersion, $f_c=6.5$ kHz, no noise).

Bandwidth (kHz)	M	Inverted v_s (m/s)	Error (%)	Iterations
0.2	2	2082.3	2.5	4
0.6	4	2082.6	2.5	5
1.0	6	2083.0	2.5	5
1.4	8	2083.7	2.5	6
1.8	10	2084.4	2.6	6
2.2	12	2085.3	2.6	6
2.6	14	2086.3	2.7	6
3.0	16	2087.3	2.7	5

TABLE III. Effect of increasing bandwidth on the conditioning of the problem of simultaneous inversion for formation shear and compressional wave velocities (fast formation, flexural dispersion, $f_c = 6.5$ kHz, no noise).

Bandwidth (kHz)	M	Condition number
0.2	2	582 000
0.6	4	64 400
1.0	6	18 800
1.4	8	7780
1.8	10	3900
2.2	12	2210
2.6	14	1370
3.0	16	903

lated for brevity). Again, a high degree of nonuniqueness of the problem precludes accurate extraction of both parameters from the data when noise is present, a situation that cannot be circumvented by increasing the bandwidth. It is pointed out that this is not because of a possible shortcoming of the inversion procedure employed but rather dictated by the underlying physics of the problem.

The conclusion here is that for the fast formation one of the parameters v_s, v_p, or a can be determined from the flexural dispersion data considered. For the slow formation, only v_s can be inverted successfully due to a lack of sensitivity to the other parameters (Sec. IV A). Increasing the bandwidth does not add enough information to the data in order to overcome this problem. It primarily helps in the identification of the dispersion curve so that, e.g., outliers in the data can be removed prior to further processing, and in providing a larger number of measured points so that noise averages out. The latter, however, could also be realized by stacking repeated narrow-band measurements at one frequency.

C. Multifrequency and multimode data

Since we found that increasing the bandwidth about an intermediate frequency of the flexural dispersion does not add any significant independent information to the data (Sec. IV B), we experimented with providing the inversion algorithm with narrow-band data collected at different frequencies and combining data belonging to different dispersion curves.

1. Multifrequency data

The formation shear wave velocity constitutes the low-frequency asymptote of the flexural dispersion. Thus, it is obvious that by combining one point taken from that section of the dispersion curve where the asymptote practically has been reached [e.g., at 2 kHz in Fig. 2(a)] and one point at an intermediate frequency (such as 6.5 kHz for the fast formation), a simultaneous inversion for v_s and one other parameter to which the second measurement is sensitive (a or v_p for the fast formation) would be possible. However, this scenario is somewhat trivial because obtaining the first point requires measuring the low-frequency part of the dispersion curve, which, if acquired with success, would yield knowledge of v_s directly. Furthermore, if a k_{zi} value is close to ω_i / v_s the cost function in Eq. (4) exhibits a kink with respect to the component of $\bar{x}$ representing the model shear wave velocity, close to the location of the pursued minimum and thus likely to interfere with the optimization. These considerations lead to the question of how far the first point can be moved up the dispersion curve (away from the knee toward higher frequencies, keeping the second measurement fixed) while the two data points can still be used to deduce v_s and a second parameter (a or v_p) simultaneously, without recreating the nonunique situation as in Sec. IV B. The outcome of this numerical experiment is that typically the transition to an underdetermined problem is a sharp one, occurring almost immediately after the data point moves away from the asymptote.

TABLE IV. Effect of increasing noise level on the simultaneous inversion for formation shear and compressional wave velocities (fast formation, flexural dispersion, $f_c = 6.5$ kHz, 16 samples with 3 kHz bandwidth).

Noise (%)	Inverted v_s (m/s)	Inverted v_p (m/s)	Error v_s (%)	Error v_p (%)	Iterations
0.0	2032.0	3658.0	0.0	0.0	7
0.5	2023.7	3784.1	−0.4	3.5	7
1.0	2227.9	2969.8	9.6	−18.8	7
1.5	2981.3	3057.0	46.7	−16.4	18

TABLE V. Effect of increasing noise level on the inversion for formation shear wave velocity with −10% error in formation compressional wave velocity (fast formation, flexural dispersion, $f_c = 6.5$ kHz, 16 samples with 3 kHz bandwidth).

Noise (%)	Inverted v_s (m/s)	Error (%)	Iterations
0.0	2087.3	2.7	5
1.0	2094.3	3.1	5
2.0	2093.4	3.0	5
3.0	2084.3	2.6	4
4.0	2067.3	1.7	4
5.0	2042.5	0.5	4

2. Multimode data

Since typical logging tools have both monopole and dipole sources available, it is reasonable to combine flexural and Stoneley mode dispersion data and to reconsider the possibility of simultaneous inversion for more than one parameter. The capability of processing such multimode data without any modifications is a major advantage of the inversion scheme introduced in Sec. II B. For the borehole problem with proper modes, on input each data point (ω_i, k_{zi}) has to be flagged only with respect to the azimuthal mode order n so that the determinant can be evaluated accordingly; no further labeling of the data is necessary.

As an example, consider the fast formation and simultaneous inversion for formation shear wave velocity v_s and borehole radius a. The combination of one point from the Stoneley mode at 5 kHz and one point from the flexural mode at 6 kHz leads to a minimum of the cost function that is much better defined than in the case of using flexural data only. However, when using the same data for the simulta-

neous inversion of v_s and formation compressional wave velocity v_p no improvement results and a long, shallow minimum of the cost function is encountered.

By providing several samples of the Stoneley and flexural dispersion curves from around 5 and 6 kHz, respectively, and dividing *M*, the number of data points, unevenly between the two modes, it is possible to change their relative importance in the inversion and influence the shape of the minimum. If the same number of data points is provided from both dispersion curves, however, the minimum is found to be similar to the case $M=2$, once again showing the weak influence of data bandwidth on the information content of the data.

For the slow formation, we found that combining Stoneley and flexural dispersion data does not help in improving the conditioning of either of the two simultaneous inversion problems (e.g., when using two points located both at 2 kHz).

V. SUMMARY AND CONCLUSIONS

We have described the problem of inverting guided-wave dispersion data with the goal of inferring unknown waveguide constitutive and geometric parameters. The forward problem of obtaining modal dispersion curves is considered first. The determinant of the system matrix of the homogeneous linear system of equations that follows from matching the appropriate boundary conditions is a function of axial wave number, angular frequency, and a waveguide parameter vector. The discrete zeros of this determinant or roots of the guidance condition correspond to different modes. The corresponding dispersion curves, giving the associated axial wave number as a function of frequency and obtained by a numerical mode tracking procedure, depend on the waveguide parameter vector. Thus, for the inverse problem, an estimate of the waveguide parameter vector from noisy samples of one or more dispersion curves can, in principle, be obtained with the help of a curve fitting procedure. However, this scheme has several disadvantages, largely caused by the nesting of the iterations of mode tracking and nonlinear optimization. Therefore, an inversion methodology is proposed that eliminates the inner of the two iterations by formulating the cost function to be minimized directly in terms of the determinant of the boundary conditions, allowing a rapid, iteration-free evaluation of the cost function. As an example, the sonic guidance condition for a fluid-filled borehole is considered. In the formulation, care is exercised in avoiding parasitic zeros and infinite values that are likely to cause problems in numerical forward and inverse dispersion analyses. The high-frequency limit, the guidance condition of the Scholte wave, is obtained as a check. For the examples of a fast and a slow rock formation, the inversion of formation shear wave velocity, a quantity of great practical interest, with an error of the assumed model in borehole radius and formation compressional wave velocity is investigated numerically. Results are given for both the flexural and the Stoneley mode (the lowest-order dipole and monopole modes, respectively), providing useful insights into the nonlinear dependence of inverted formation shear wave velocity on the model errors. The frequency-dependent sensitivities of inverted shear wave velocity to the model errors are also displayed. These sensitivities are not to be confused with the elements of the Jacobian in the Gauss–Newton method used to perform the inversion via nonlinear optimization. The simultaneous inversion of formation shear and compressional wave velocities from flexural dispersion data in the fast formation is studied next, varying the bandwidth of the provided data and the level of contaminating noise. It is found that the simultaneous inversion problem is difficult due to the intrinsic insensitivity of the data with respect to the formation compressional wave velocity. Increasing the bandwidth is beneficial, but less than desirable. The inversion for formation shear wave velocity only is found to be robust in the presence of noise and is little dependent on bandwidth. A numerical experiment using multifrequency data for simultaneous inversion is described and assessed. Combining data from different dispersions—in the example the flexural and Stoneley dispersions in the fast formation—can enable the simultaneous inversion of formation shear wave velocity and borehole radius.

[1] J. A. Kong, *Electromagnetic Wave Theory* (EMW Publishing, Cambridge, 2000).
[2] S. W. Lang, A. L. Kurkjian, J. H. McClellan, C. F. Morris, and T. W. Parks, ''Estimating slowness dispersion from arrays of sonic logging waveforms,'' Geophysics **52**, 530–544 (1987).
[3] M. P. Ekstrom, ''Dispersion estimation from borehole acoustic arrays using a modified matrix pencil algorithm,'' in *Conference Record of the 29th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, 31 October 1995, Vol. 1, pp. 449–453, IEEE Comp. Soc. Press, Los Alamitos, 1996.
[4] J. L. Stevens and S. M. Day, ''Shear velocity logging in slow formations using the Stoneley wave,'' Geophysics **51**, 137–147 (1986).
[5] B. K. Sinha, ''Sensitivity and inversion of borehole flexural dispersions for formation parameters,'' Geophys. J. Int. **128**, 84–96 (1997).
[6] C. H. Cheng, M. N. Toksöz, and M. E. Willis, ''Determination of *in situ* attenuation from full waveform acoustic logs,'' J. Geophys. Res. **87**, 5477–5484 (1982).
[7] K. J. Ellefsen, C. H. Cheng, and M. N. Toksöz, ''Applications of perturbation theory to acoustic logging,'' J. Geophys. Res. **96**, 537–549 (1991).
[8] B. K. Sinha and S. Kostek, ''Stress-induced azimuthal anisotropy in borehole flexural waves,'' Geophysics **61**, 1899–1907 (1996).
[9] B. K. Sinha and K. W. Winkler, ''Formation nonlinear constants from sonic measurements at two borehole pressures,'' Geophysics **64**, 1890–1900 (1999).
[10] B. K. Sinha, M. R. Kane, and B. Frignet, ''Dipole dispersion crossover and sonic logs in a limestone reservoir,'' Geophysics **65**, 390–407 (2000).
[11] C. V. Kimball, ''Shear slowness measurement by dispersive processing of the borehole flexural mode,'' Geophysics **63**, 337–344 (1998).
[12] C. V. Kimball and D. J. Scheibner, ''Error bars for sonic slowness measurements,'' Geophysics **63**, 345–353 (1998).
[13] H. Braunisch, T. M. Habashy, B. K. Sinha, and J. Pabon, ''Inversion of borehole dispersions for formation elastic moduli,'' in *Proceedings of the IEEE International Ultrasonics Symposium*, San Juan, 22–25 October 2000, Vol. 1, pp. 551–556.
[14] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in* FORTRAN 77: *The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
[15] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization* (Academic, New York, 1981).
[16] K. F. Graff, *Wave Motion in Elastic Solids* (Dover, New York, 1991).
[17] B. K. Sinha and S. Asvadurov, ''Higher-order modes in a fluid-filled borehole,'' in *Proceedings of the IEEE International Ultrasonics Symposium*, Sendai, 5–8 October 1998, Vol. 2, pp. 1115–1120.
[18] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (National Bureau of Standards, Washington, DC, 1964).

[19] A. N. Norris and B. K. Sinha, ''The speed of a wave along a fluid/solid interface in the presence of anisotropy and prestress,'' J. Acoust. Soc. Am. **98**, 1147–1154 (1995).
[20] C.-H. Hsu and B. K. Sinha, ''Mandrel effects on the dipole flexural mode in a borehole,'' J. Acoust. Soc. Am. **104**, 2025–2039 (1998).
[21] M. A. Biot, ''Propagation of elastic waves in a cylindrical bore containing a fluid,'' J. Appl. Phys. **23**, 997–1005 (1952).
[22] H. Lamb, ''On the velocity of sound in a tube as affected by the elasticity of the walls,'' *Memoirs and Proceedings of the Manchester Literary and Philosophical Society*, July 1898, Vol. 42, No. 9, pp. 1–16.

Representation of concurrent acoustic objects in primary auditory cortex

Benjamin J. Dyson[a] and Claude Alain
Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, Ontario M6A 2E1, Canada

(Received 11 July 2003; revised 3 October 2003; accepted 20 October 2003)

Auditory scene analysis involves the simultaneous grouping and parsing of acoustic data into separate mental representations (i.e., objects). Over two experiments, we examined the sequence of neural processes underlying concurrent sound segregation by means of recording of human middle latency auditory evoked responses. Participants were presented with complex sounds comprising several harmonics, one of which could be mistuned such that it was not an integer multiple of the fundamental frequency. In both experiments, Na (approximately 22 ms) and Pa (approximately 32 ms) waves were reliably generated for all classes of stimuli. For stimuli with a fundamental frequency of 200 Hz, the mean Pa amplitude was significantly larger when the third harmonic was mistuned by 16% of its original value, relative to when it was tuned. The enhanced Pa amplitude was related to an increased likelihood in reporting the presence of concurrent auditory objects. Our results are consistent with a low-level stage of auditory scene analysis in which acoustic properties such as mistuning act as preattentive segregation cues that can subsequently lead to the perception of multiple auditory objects. [DOI: 10.1121/1.1631945]

PACS numbers: 43.64.Ri, 43.64.Qh, 43.66.Lj [WPS] Pages: 280–288

I. INTRODUCTION

One condition necessary for successful interaction with the environment is the identification of numerous sound sources that traditionally overlap in both space and time. This general problem of auditory scene analysis, and the more specific task of detailing those acoustic cues that are critical for perception of distinct auditory objects, have received considerable attention over the last two decades [for reviews, see Hartmann (1988), Bregman (1990), and Kubovy and Van Valkenburg (2001)]. For example, variations in onset/offset synchrony and spatial locale can influence the perception of singular or multiple auditory objects within an acoustic scene. In addition to these temporal and spatial cues, complex sounds consisting of multiple tonal elements can also be grouped and segregated on the basis of frequency periodicity (i.e., harmonic series). Previous research has shown that the likelihood of concurrently hearing more than one auditory object increases as the amount of mistuning increases for lower tonal components of a harmonic series, resulting in the ''pop out'' of lower mistuned partials from the complex (Moore *et al.*, 1986; Chalikia and Bregman, 1989).

Evidence for the use of mistuning as an early concurrent sound segregation cue is suggested by animal studies. Results from single cell recordings in mammals indicate that inharmonicity is already extracted and represented within the firing patterns of neurons in the auditory nerve for guinea pigs (Palmer, 1990), and in the inferior colliculus for chinchillas (Sinex *et al.*, 2002). These results are consistent with the notion that concurrent sound segregation based on frequency periodicity is carried out very early on in the auditory pathway.

In humans, the neural events associated with the perception of multiple auditory objects as a result of mistuning have been investigated using event-related brain potentials (ERPs). The ERPs elicited by auditory stimuli are divided into three groups on the basis of their latencies: (*i*) auditory brainstem responses occurring from stimulus onset to 10 ms poststimulus; (*ii*) middle latency auditory evoked potentials observed between 10 and 50 ms poststimulus thought to index activity in the thalamus and primary auditory cortex; and (*iii*) long latency auditory evoked potentials arising after 50 ms poststimulus as a result of activation in primary and associative auditory cortex (Picton and Fitzgerald, 1983; Ballantyne, 1990). To date, the earliest reported change in ERPs for mistuned harmonics is an increase in negativity observed between 140 and 240 ms after sound onset (Alain *et al.*, 2002). Source analyses suggest this effect originates from the superior temporal gyrus, although it is unclear whether this generator is located in primary or associative auditory cortices (Alain *et al.*, 2001).

Upon this reading, the time course of neural events associated with the detection of mistuning appears to be relatively late in humans, compared to the animal literature. However, this apparent discrepancy may be due to the fact that previous electrophysiological research in humans has tended to focus on later auditory evoked responses. Therefore, the rationale for measuring middle latency auditory evoked responses (MLAERs) for tuned and mistuned stimuli in humans was twofold. First, demonstrating the modulation of MLAERs as a result of mistuning would provide an important bridge between animal and human research. Second, the observation of MLAERs for mistuning would implicate a neural generator in the primary auditory cortex (Scherg and von Cramon, 1986), thereby supporting the notion that inharmonicity is a critical and basic cue employed in auditory scene analysis.

[a]Electronic mail: bdyson@rotman-baycrest.on.ca

II. EXPERIMENT 1

A. Method

1. Participants

Informed consent for 12 participants (between 20 and 37 years of age; six men) was obtained before the experiment was conducted. One individual was replaced with an additional participant due to the presence of 60-Hz noise within the ERP recording. All individuals had previously been assessed with respect to hearing thresholds, and were found to be within the normal range. All volunteers received an honorarium for their participation.

2. Stimuli and task

Five complex sounds were created by combining ten pure tones of equal intensity. All stimuli had a fundamental frequency of 200 Hz and a duration of 40 ms with 2.5 ms linear onset and offset ramps. The five stimuli varied with respect to the upward tuning of the third harmonic (0%, 2%, 4%, 8%, and 16%), while all other components remained harmonically related to the fundamental frequency. The sounds were digitally generated with a sampling rate of 24 kHz and passed through an anti-aliasing filter at 10 kHz, using a 16-bit Tucker Davis System converter (RP2; Alachua, FL). All stimuli were presented binaurally over Telephonics headphones (TDH-50D; Farmingdale, NY) at 85 dB SPL, as measured by an artificial ear and sound level meter (Larson Davis System 824; Provo, UT).

Participants listened to six blocks of trials, with each block consisting of 3000 trials. The five stimuli were presented over 600 trials each, with trial order randomized within each block. The interstimulus interval (ISI) varied randomly from 60 to 260 ms in 10-ms steps. ISI was random in order to reduce the systematic contribution of long latency responses from the previous trial on the prestimulus interval. Stimulus presentation took place in a sound-attenuating chamber and each block was initiated by the experimenter. During stimulus presentation, participants sat in a comfortable chair and were asked to watch a silent film with subtitles while ignoring the auditory stimuli. After the completion of all six blocks, participants were thanked for their time and debriefed as to the nature and purpose of the investigation.

3. Electrophysiological recording

The electrophysiological response of each participant was digitized continuously with a bandpass of 1–500 Hz and sampling rate of 2000 Hz from an array of 64 electrodes using NeuroScan SynAmps (El Paso, TX), and stored for later off-line analysis. Vertical and horizontal eye movements were recorded using electrodes at the outer canthi and at the inferior and superior orbits. During recording, all electrodes were referenced to Cz, while for data analysis they were re-referenced to an average reference.

Average electrical activity for each stimulus type was collapsed across block according to an epoch consisting of 10 ms of prestimulus baseline activity and 100 ms of poststimulus activity. Epochs with peak-to-peak deflections exceeding $\pm 100\ \mu$V at electrodes not adjacent to the eyes were excluded from the average. In addition, ocular artifacts were corrected using Brain Electrical Source Analysis (BESA; Germany) software (Lins *et al.*, 1993a, b; Berg and Scherg, 1994; Picton *et al.*, 2000). The waveforms underwent a 10-Hz high-pass and 200-Hz low-pass digital filtering.

4. Data analysis

The effects of mistuning on the latency and amplitude of the MLAERs at fronto-central sites was examined using a two-way repeated measured ANOVA with the factor of mistuning having five levels (0%, 2%, 4%, 8%, and 16%) and the factor of electrode having six levels (F1, Fz, F2, FC1, FCz, and FC2). In terms of inferential statistics, only the Na and Pa deflections were analyzed in further detail, given that the preceding MLAERs of No and Po and the subsequent MLAERs of Nb and Pb are thought to be largely unreliable in nature (Amenedo and Diaz, 1998; Beiter and Hogan, 1973; Thornton *et al.*, 1977; Woods and Clayworth, 1995). Peak latencies for Na and Pa were quantified according to the time ranges reported in Woods *et al.* (1993, 1995). Peak latency and amplitude for Na were defined as the maximum negativity between 15 and 25 ms after stimulus onset, while peak latency and amplitude for Pa were defined as the maximum positivity between 25 and 36 ms after stimulus onset. Finally, mean Na and Pa amplitude was calculated over a 2-ms time window, defined relative to 1 ms either side of the peak latency value. This value was traditionally based on the group mean of peak latency over all levels of mistuning and electrode, unless analyses revealed significant differences in peak latency across these two factors. In such cases, variation in peak latency for individual levels of mistuning and/or electrode were taken into account and separate mean amplitude time windows were calculated accordingly. The original degrees of freedom are reported, although significance levels were based upon reduced degrees of freedom as determined by the Greenhouse-Geisser epsilon. Any main effects of mistuning were further explored using Tukey's HSD test.

B. Results

All stimuli generated large Na and Pa deflections that were maximal at fronto-central sites and inverted in polarity at inferior temporal-parietal sites, consistent with a neural generator in the auditory cortex along the Sylvian fissure. Figure 1 shows the group mean MLAERs over the six fronto-central electrode sites for 0%, 4%, and 16% mistuning. The Na peak latency and peak amplitude did not differ as a function of mistuning, $F(4,44) = 1.15$ and 1.55, respectively, $p > 0.215$. Mistuning also failed to have a statistically significant main effect on Na mean amplitude: $F(4,44) = 1.25$, $p = 0.306$. Upon inspection of Fig. 1, the lack of significance for the Na wave is surprising. Na amplitude appears to be modulated by the degree of mistuning within the stimulus, particularly at electrodes F1, Fz, and F2. Indeed, a statistically significant mistuning$\times$electrode interaction was observed for Na peak amplitude: $F(20,220) = 2.45$, $p = 0.050$, $\varepsilon = 0.233$. Further analysis of Na peak amplitude over these three sites however again revealed a nonsignificant main effect of mistuning: $F(4,44) = 2.05$, $p = 0.116$.

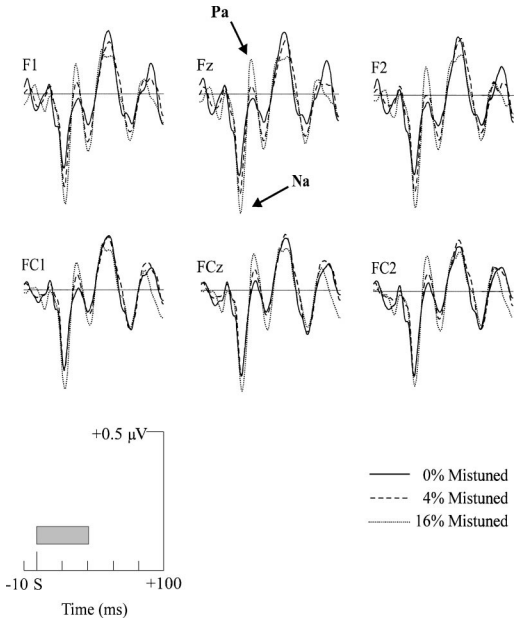


FIG. 1. Group mean middle latency auditory evoked responses (MLAERs) from six fronto-central electrode sites (F1, Fz, F2, FC1, FCz, and FC2) as a function of mistuning (0%, 4%, and, 16%) in experiment 1. The gray rectangle indicates the duration of the stimulus.

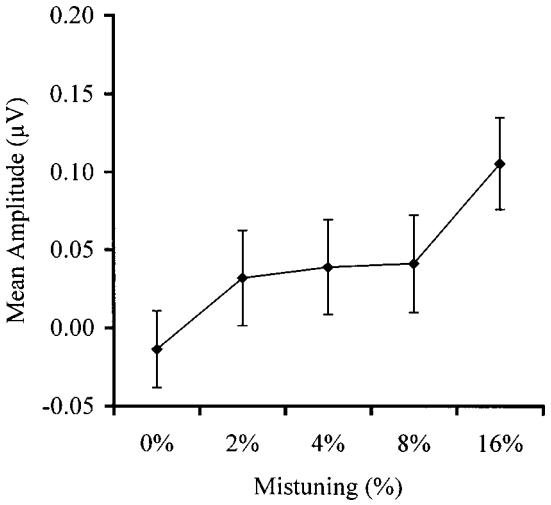


FIG. 2. Group mean Pa mean amplitude collapsed over electrode site, as a function of mistuning in experiment 1. Error bars represent standard error.

Clearly, the variability of the Na component in experiment 1 precludes a definitive modulation with respect to mistuning.

In the case of Pa, the ANOVA revealed a statistically significant main effect of mistuning on peak latency: $F(4,44)=3.14$, $p=0.040$, $\varepsilon=0.734$. Tukey's HSD test revealed that the peak latency for the 2% mistuned stimulus occurred significantly later than the 0% and 8% mistuned stimulus ($p<0.05$), with all other pairwise comparisons failing to reach statistical significance ($p>0.05$). One possible explanation for this result is that the 2% stimulus is an ambiguous stimulus that lies on the threshold of mistuning detection (cf. Moore *et al.*, 1986) and therefore evokes a later peak latency for Pa. There was also a main effect of mistuning across fronto-central sites for both peak and mean amplitude, $F(4,44)=3.63$ and 6.40, $p<0.05$ in both cases, $\varepsilon=0.561$ and 0.618, respectively. Tukey's HSD test revealed that the 16% mistuned stimulus generated significantly greater Pa peak and mean amplitude than the 0% and 2% mistuned stimuli ($p<0.05$), with all other pairwise comparisons failing to reach significance ($p>0.05$). Figure 2 summarizes this change in mean Pa amplitude as a function of mistuning.

The relation between Pa enhancement and the participants' phenomenological experience was examined in 8 of the original 12 participants, who took part in an additional behavioral study [after Alain *et al.* (2001)]. Each of the five complex sounds used in experiment 1 were presented 40 times in a random order over a single block of 200 trials, with a response-stimulus interval of 1000 ms. Participants were required to respond on each trial whether the stimulus was more like one or two sounds, by pressing separate buttons on a response box. Performance feedback was not provided. Figure 3(a) shows the likelihood of reporting hearing two auditory objects as a function of mistuning. There was a main effect of stimulus type: $F(4,28)=91.71$, $p<0.001$. Tukey's HSD test showed all pairwise comparisons to be significant ($p<0.05$), apart from the comparison between 0% and 2%, and between 8% and 16% (both $p>0.05$). To ensure that these eight participants were representative of the overall sample with respect to their middle latency responses, Pa mean amplitude from this subset of individuals was reanalyzed [see Fig. 3(b)], replicating the significant main effect of mistuning: $F(4,28)=5.83$, $p=0.007$, $\varepsilon=0.654$. Tukey's HSD test revealed that 16% mistuning produced significantly greater Pa mean amplitude than all other categories of stimuli ($p<0.05$), with all other pairwise comparisons failing to reach significance ($p>0.05$). The relationship between the probability of detecting two sounds and Pa mean amplitude is shown in Fig. 3(c), with each point in the scatterplot representing the response of a single participant at a specific level of mistuning. In order to quantify this relationship, individual correlations were calculated for each of the eight participants and the group mean correlation was entered into a one-sample *t*-test [after Alain *et al.* (2001)]. This showed a significant positive relationship between the probability of responding two sounds and Pa mean amplitude: $r=0.49$, $t(7)=2.58$, $p=0.037$, indicating that this relatively early neural correlate of concurrent sound segregation was predictive of later perceptual judgments.

Experiment 1 provided initial evidence to suggest that mistuning is represented early on in the auditory pathway in humans as indexed by modulation of the Pa wave. Moreover, this wave was also related to the likelihood of hearing concurrent multiple auditory objects. However, given the novelty of this finding and that a number of previous researchers have questioned the reliability of middle latency responses (Beiter and Hogan, 1973; Thornton *et al.*, 1977; Amenedo and Diaz, 1998; Woods and Clayworth, 1985), we thought it necessary to replicate the basic effect of mistuning on Pa mean amplitude. Experiment 2 was designed with an aim of focusing directly upon the mistuning effect by employing both a subset of stimuli and electrodes used in experiment 1. Specifically, activity was measured at Fz and Cz since the

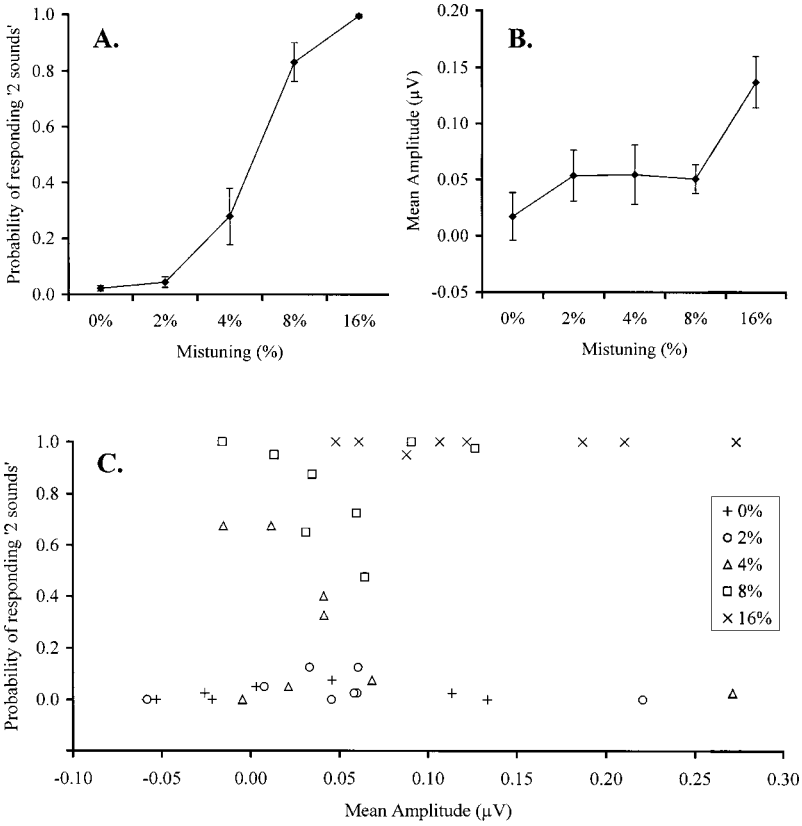


FIG. 3. The effect of mistuning in experiment 1 for a subset of eight participants plotted against: (a) the probability of perceiving two objects and (b) Pa mean amplitude collapsed over electrode site. In both cases, error bars represent standard error. (c) A scatterplot of the relationship between the probability of perceiving two objects and Pa mean amplitude for each of the eight participants over five levels of mistuning (0%, 2%, 4%, 8%, and 16%). The *y*-axis has been artificially raised in order to provide a clearer representation of the data.

effect of mistuning was maximal at fronto-central sites in experiment 1.

III. EXPERIMENT 2

A. Method

1. Participants

Informed consent for 12 participants (between 19 and 35 years of age; 4 men) was obtained before the experiment was conducted. Two individuals were replaced with additional participants in experiment 2: one had to be substituted due to the presence of an electrical artifact within the data, while a second was replaced after an error in the intensity level of stimulus presentation was revealed during the session. All individuals were assessed with respect to hearing thresholds, and were found to be within the normal range for frequencies between 250 to 4000 Hz (both ears). All volunteers received an honorarium for their participation.

2. Stimuli and task

Four stimuli were created for experiment 2. Once again, they were 40-ms complex sounds comprising ten pure tone elements of equal intensity. The stimuli differed with respect to fundamental frequency (200 or 400 Hz) and the degree of upward mistuning for the third harmonic (0% or 16%), with all other components remaining harmonically related to the fundamental. In an attempt to improve the signal-to-noise ratio, the number of trials within each block was increased. Participants now listened to six blocks of trials each consisting of 3600 stimulus presentations, with each of the four stimuli being presented for 900 trials in a random order. In addition, onset and offset ramps were reduced from 2.5 to 1.5 ms (cf. Lane *et al.*, 1971; Skinner and Antinoro, 1971). All sounds were calibrated to 85 dB SPL and played over Eartone 3A insert earphones (Indianapolis, IN) while participants watched a silent movie in a sound-attenuating chamber. All other aspects of the stimuli and task were identical to experiment 1.

3. Electrophysiological recording and data analysis

Only a subset of electrodes from experiment 1 were employed in experiment 2. Specifically, electrical activity from Fz, Cz, and TP10 were recorded, with TP9 acting as the on-line reference and a collarbone electrode acting as the ground. Eye movements and other artifacts were excluded by adopting a stringent ± 50 μV epoch rejection criterion. During off-line data analysis, all electrodes were re-referenced to a linked mastoid (TP9 and TP10) reference. With respect to all other parameters, the recording and analysis of event-related potentials in experiment 2 were identical to experiment 1. Group mean MLAERs were analyzed using a three-way repeated measures ANOVA with the factors of mistuning (0% and 16%), fundamental frequency (200 and 400 Hz), and site (Fz and Cz).

B. Results

Figure 4 shows the group mean MLAERs over Fz and Cz for the four different combinations of fundamental frequency and mistuning in experiment 2. Na peak amplitude revealed a significant main effect of fundamental frequency: $F(1,11) = 11.52$, $p = 0.006$, but not for mistuning: $F(1,11) = 1.93$, $p = 0.192$. Both fundamental frequency and mistuning failed to show any significant main effects for peak la-

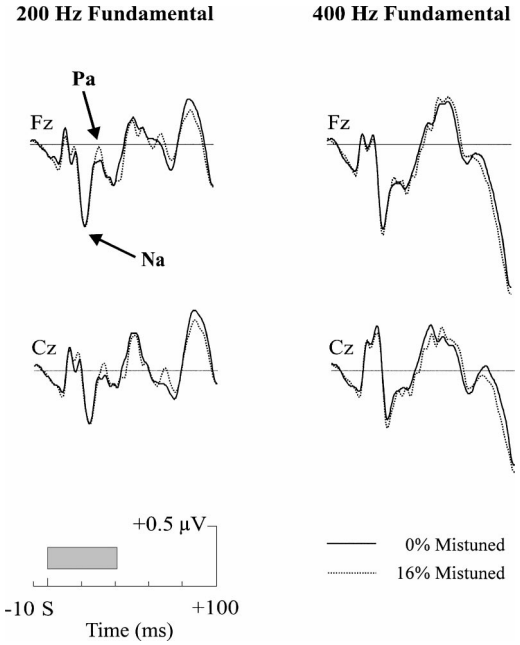


FIG. 4. Group mean middle latency auditory evoked responses (MLAERs) from two electrodes (Fz and Cz) as a function of fundamental frequency (200 and 400 Hz) and mistuning (0% and 16%) in experiment 2. The gray rectangle indicates the duration of the stimulus.

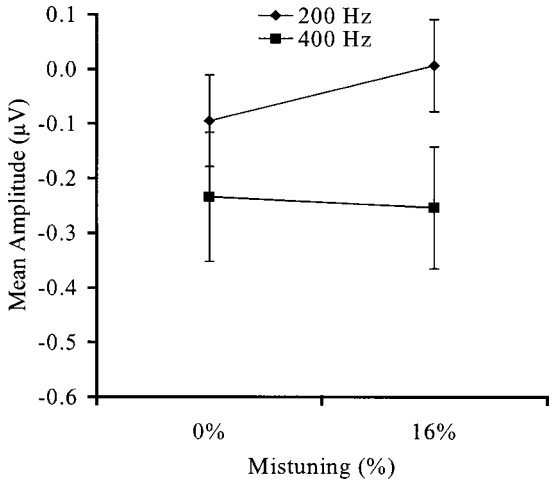


FIG. 5. Group mean Pa mean amplitude collapsed over electrode site, as a function of mistuning and fundamental frequency in experiment 2. Error bars represent standard error.

tency: $F(1,11)=0.25$, $p=0.625$ and $F(1,11)=0.44$, $p=0.521$ respectively, or mean amplitude: $F(1,11)=0.06$, $p=0.814$ and $F(1,11)=0.57$, $p=0.467$, respectively.

Pa peak amplitude revealed a significant main effect of fundamental frequency only: $F(1,11)=5.55$, $p=0.038$, with the main effect of mistuning failing to reach significance: $F(1,11)=2.50$, $p=0.142$. Peak latency demonstrated no significant main effects with respect to fundamental frequency or mistuning: $F(1,11)=1.89$, $p=0.196$, and $F(1,11)=0.42$, $p=0.531$, respectively. Mean amplitude, however, showed a significant main effect of fundamental frequency: $F(1,11)=12.61$, $p=0.005$; a nonsignificant main effect of mistuning: $F(1,11)=1.01$, $p=0.336$; and a statistically significant interaction between fundamental frequency and mistuning: $F(1,11)=6.75$, $p=0.025$ (see Fig. 5). Tukey's HSD test revealed that all pairwise comparisons were statistically different from one another ($p<0.05$) apart from the mean Pa amplitude generated for the tuned (0%) and mistuned (16%) 400-Hz stimuli. Therefore, experiment 2 replicates the main finding of experiment 1, namely, 16% mistuning of the third harmonic in a 200-Hz complex stimulus significantly increases the mean Pa amplitude relative to a tuned 200-Hz complex stimulus.

Once again, the link between Pa amplitude and the perception of multiple auditory objects was probed by calling 8 of the original 12 participants from experiment 2 back into the laboratory to complete a behavioral study. The details of the task were identical to the follow-up behavioral task described in experiment 1, apart from the single block of 200 trials was now composed of four stimuli (200 Hz 0%, 200 Hz 16%, 400 Hz 0%, and, 400 Hz 16%) presented for 50 trials each in a random order. As before, participants were asked to indicate without feedback whether the stimulus at each trial sounded more like one sound (i.e., a buzz) or two sounds [i.e., a buzz plus another sound with a pure tone quality; see Alain *et al.*, (2002)]. The behavioral data was analyzed according to a two-way repeated measures ANOVA with the factor of mistuning (0% and 16%) and fundamental frequency (200 and 400 Hz). The results of the ANOVA yielded statistically significant main effects for fundamental frequency: $F(1,7)=5.82$, $p=0.047$, and mistuning: $F(1,7)=37.59$, $p<0.001$, in addition to a statistically significant interaction between mistuning and fundamental frequency: $F(1,7)=6.83$, $p=0.035$. This interaction was decomposed using Tukey's HSD test and is summarized in Fig. 6(a). This *posthoc* test showed all pairwise comparisons to be significant ($p<0.05$), apart from the comparison between 200 Hz

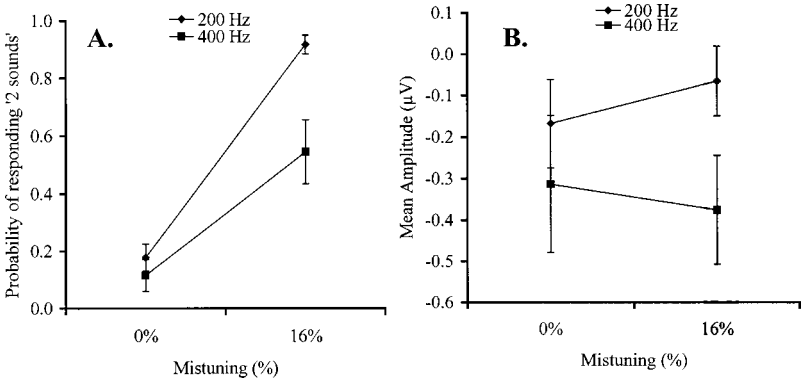


FIG. 6. The effect of mistuning in experiment 2 for a subset of eight participants plotted against: (a) the probability of perceiving two objects and (b) Pa mean amplitude collapsed over electrode site. Error bars represent standard error.

0% mistuning and 400 Hz 0% mistuning ($p>0.05$). The behavioral data demonstrate that while the presence of mistuning increases the likelihood of hearing two objects, mistuning the third harmonic of a 200-Hz complex stimulus by 16% is more likely to lead to the perception of multiple auditory objects, relative to a 400-Hz complex tone whose third harmonic is mistuned to the same degree. This accords well with the MLAER data for experiment 2 in which mistuning produced a significant increase in Pa amplitude for the 200-Hz case only.

As before, to ensure that these eight participants were representative of the sample with respect to their MLAERs, Pa mean amplitude was reanalyzed with this subset of individuals and the mistuning×fundamental frequency interaction was replicated: $F(1,7)=11.41$, $p=0.012$ [see Fig. 6(b)]. Tukey's HSD test revealed that tuned and mistuned 400-Hz stimuli produced significantly lower Pa mean amplitude than tuned and mistuned 200-Hz stimuli ($p<0.05$), with all other pairwise comparisons failing to reach significance at the 0.05 level. The critical comparison between 200 Hz 0% and 200 Hz 16% mistuned, however, approached statistical significance ($p=0.077$).

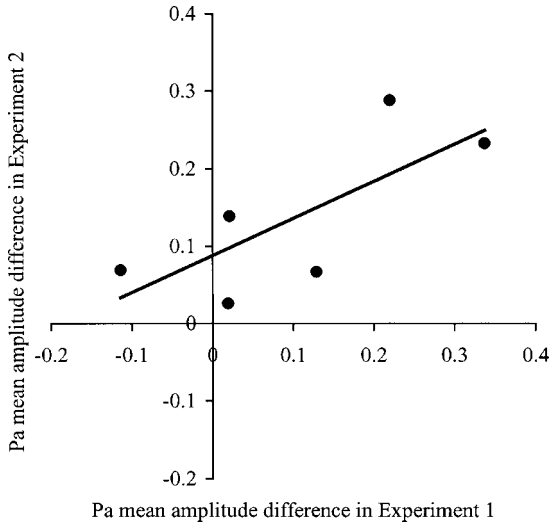


FIG. 7. The difference in Pa mean amplitude between the 200-Hz mistuned (i.e., 16%) and tuned (i.e., 0%) stimulus as measured within the same six participants during experiment 1 and, then, approximately 3–4 months later, during experiment 2.

C. Within subject reliability of MLAERs

One consequence of the subject recruitment procedure was that half of the individuals who took part in experiment 1 also completed experiment 2. Therefore, as a final assessment of the reliability of MLAERs within the current experimental series, the difference in MLAERs between a tuned (i.e., 0%) and mistuned (i.e., 16%) 200-Hz stimulus was compared during experiment 1 and then, approximately 3–4 months later, during experiment 2. If the effect of mistuning was reliable within participant across time, then a positive correlation should obtain between these two measures. In order to increase the validity of the comparison between experiments 1 and 2, data from experiment 1 were recalculated according to a linked mastoid reference and the average was based on collapsed Fz and Cz sites only. As graphically represented in Fig. 7, we found a significant positive correlation between Pa mean amplitude difference (200 Hz 16% minus 200 Hz 0%) recorded across experiments 1 and 2 for this subset of participants: $r=0.74$, $p=0.045$ (one-tailed), thereby supporting the notion that Pa responses to mistuning may be reliable over time within the same individuals.

IV. DISCUSSION

The complex sounds used in experiments 1 and 2 generated reliable Na and Pa responses, with the timing of these components (approximately 22 and 32 ms, respectively) consistent with those reported by previous research employing pure tone stimuli (e.g., Lane *et al.*, 1971; Thornton *et al.*, 1977; Woods *et al.*, 1993, 1995). Moreover, the first reliable effect of mistuning observed within MLAERs was the modulation of Pa mean amplitude. This effect of inharmonicity on MLAERs precedes previously reported modulation of human ERP as a function of mistuning by around 100 ms (cf. Alain *et al.*, 2001). We believe that our finding regarding the early registering of mistuning information results in (*i*) a narrowing of focus with respect to the localization of the neural generators responsible for the detection of mistuning, (*ii*) the consideration of mistuning as a preattentively available acoustic property, and (*iii*) the treatment of inharmonicity as a fundamental cue in concurrent sound segregation during auditory scene analysis.

First, the latency at which inharmonicity is registered within the current experimental series severely limits the possible locations for the underlying neural generators that produce the effect. With mistuning registered at approximately 32 ms after stimulus onset, the data strongly point to an encoding of mistuning which occurs very early on in the auditory pathway. Converging evidence from dipole source modeling (Scherg and von Cramon, 1986), intracebral recording in human (Liégois-Chauvel *et al.*, 1991), magnetoencephaloghrapy (Pantev *et al.*, 1995), and animal work (McGee *et al.*, 1992) suggests that the Pa wave is generated in the primary auditory cortex [for a review, see Eggermont and Ponton (2002)]. Additionally, the fact that the topography for the Pa elicited by the tuned and mistuned stimuli did not significantly differ from one another suggests that the enhanced amplitude reflects a modulation of the Pa generator rather than the contribution from another source which was active during the Pa interval. The observation of Pa modulation at fronto-central sites in the present study is also consistent with localization of a neural source in the Sylvian fissure along the supratemporal plane (Alain *et al.*, 2001). Therefore, given that the localization of neural generators is currently one of the weaker aspects of ERP research (Eggermont and Ponton, 2002), the consideration of earlier components such as MLAERs drive the researcher towards a limited number of locations in which the neural origin of the effect may be identified.

Second, previous research suggests that concurrent sound segregation as a result of the detection of mistuning may depend on low-level mechanisms that operate independently of attention (Bregman 1990; Alain and Izenberg,

2003). Our current findings are consistent with this proposal for the following reasons. Given the early latency of the current effect, the likely origin of the registration of mistuning information is within the primary auditory cortex. Consequently, the extent to which such information is sensitive to top-down influences such as attention is negligible in the present study, relative to other acoustic information which is available higher up the auditory pathway. Although Woldorff *et al.* (1993) reported an auditory selective attention effect as early as 20 ms after sound onset, the experimental conditions under which this effect was observed were critically different from the paradigm employed in the current study. Specifically, the effect shown by Woldorff *et al.* (1993) was derived from an intramodal attention task in which participants paid attention to one stream of sounds in order to detect rare and difficult to discriminate targets, while simultaneously ignoring a second and spatially distinct stream of sounds. Therefore, if high attentional load is critical to the derivation of such early effects, it seems unlikely that similar influences would have arisen here given the lack of substantive attentional demands in the current experimental design. Moreover, the significant effect of mistuning observed across experiments 1 and 2 was produced by sounds that were essentially presented outside the focus of attention. That is, overt responses to the auditory stimuli were not required and participants were in fact instructed to ignore the sounds during the viewing of the film. Indeed, even if it were to be argued that the presentation of an engaging visual stimulus such as a feature film increased the overall attentional demands of the current paradigm, there would be no *a priori* reason to assume that such a salient stimulus should selectively enhance the Pa amplitude to complex sounds with a 16% mistuned harmonic, relative to a similarly complex but harmonically related stimulus. Therefore, we propose that the detection of inharmonicity is a preattentive characteristic of auditory stimulus processing.

While attention does not appear to play a critical role in the current paradigm, other explanations regarding the difference in Pa mean amplitude between tuned and mistuned stimuli were also considered. For example, participants may be sensitive to differences in the overall acoustic energy of the stimulus, leading to Pa modulation, rather than inharmonicity *per se*. Computer simulations of the 0% (tuned) and 16% (mistuned) stimuli were generated and normalized instantaneous power (defined as signal amplitude at each time point squared) was compared between the two cases. While power differences were observed between tuned and mistuned stimuli, these differences were always characterized as larger instantaneous power in the case of the tuned stimulus. Assuming a positive relationship between instantaneous power and Pa mean amplitude, the results of the simulation show that a consideration of instantaneous power actually works against the experimental finding of increased Pa amplitude for the mistuned stimulus.

One interesting feature of experiment 2 was the relationship between the electrophysiological and behavioral results, and, more specifically, the observation that while an effect of mistuning was observed for both kinds of measure in the case of the 200-Hz stimulus, only behavioral data showed a mistuning effect for the 400-Hz stimulus. Previous studies have also noted difficulty in the detection of inharmonicity (Moore *et al.*, 1985) and a reduction in the probability of reporting the presence of two objects (Alain *et al.*, 2001) for mistuning around the fourth harmonic of a 400-Hz complex tone. A standard way of thinking about the detection of mistuning is provided by a harmonic sieve (Scheffers, 1983) or template (Lin and Hartmann, 1998) account. Here, expectations regarding the harmonic series are generated on the basis of the fundamental frequency of the stimulus in question. Inharmonic partials separate out from the rest of the harmonic series when there is a discrepancy between the expected frequency of the partial and its actual mistuned value. However, the straightforward application of this idea to the present data fails to account for the observed differences between fundamental frequency, as the actual frequency value of the 16% mistuned third harmonic for both the 200- and 400-Hz stimulus lie approximately at half cycle, assuming an equivalent width of harmonic filter for both values of fundamental frequency. That is, 696 Hz (16% mistuning of 600 Hz) is approximately halfway inbetween the expected 600- or 800-Hz partials in the case of the 200-Hz stimulus, just as 1392 Hz (16% mistuning of 1200 Hz) is approximately halfway in-between the expected 1200- or 1600-Hz partials in the case of the 400-Hz stimulus. However, a harmonic sieve account may begin to explain the difference between the 200- and 400-Hz conditions, if one assumes that there is a limitation on the number of filters that may be set up at any one time. Given that the presentation of 200- and 400-Hz stimuli was randomly intermixed in experiment 2, it is possible that participants could only create a single harmonic template per block, resulting in the reliable detection of mistuning for only a certain class of stimuli (i.e., $f0$ of 200 Hz). An alternative argument may be based on the observation that certain harmonic partials were the same for both 200- and 400-Hz stimuli (400, 800, and 1600 Hz), which may have compromised the fidelity of the 400-Hz template. However, given that an equal number of stimuli from both conditions were presented within each block, there is currently no *a priori* reason to assume that the 200-Hz stimuli should have been the exclusive class of acoustic signal sensitive to neural correlates of mistuning.

A second possibility suggested by Moore *et al.* (1985) is that, for the current sample, 1392 Hz may fall within a transitional range of frequencies in which the partial is neither detected by ''pop-out'' (associated with lower series harmonics) nor by a ''beat'' sensation (associated with higher series harmonics). As they describe for one individual who had difficulty detecting mistuning in the fourth harmonic of a 400-Hz complex tone: ''...for this particular condition the partial could not be clearly heard out from the complex, but at the same time beats could not be heard...'' (Moore *et al.*, 1985, p. 1863). The idea that the 16% mistuning of the third harmonic for a 400-Hz stimulus is perceptually ambiguous is supported by both the reduction in probability and the increase in variability of responding ''2 sounds'' in the 400-Hz mistuned condition relative to the 200-Hz mistuned condition [see Fig. 6(a)]. Although it is possible to argue that while this specific partial in the 400-Hz case fails to be reli-

ably registered by MLAERs, 16% mistuning of the third harmonic for a 400-Hz stimulus may be represented by later ERPs generated further along the auditory pathway (Alain *et al.*, 2001), which also contribute to the perception of multiple sounds. Therefore, the discrepancy between the electrophysiological and behavioral results in the 400-Hz case may be a matter of degree driven by the summed contribution of a number of neural populations, rather than a qualitative difference in the processing of 200- and 400-Hz complex stimuli.

In summary, it is worthwhile noting that the kinds of acoustic information that are represented by MLAERs are exactly those stimulus qualities which are also thought to aid sound segregation. For example, differences in pitch (Woods *et al.*, 1993; 1995), location (Woods and Clayworth, 1985), and loudness (Thornton *et al.*, 1977) have all been found to modulate MLAERs, just as these acoustic properties are also thought to assist in auditory figure-ground separation. This process of acoustic segregation provides a framework within which multiple auditory objects may be parsed from the current sound environment (Neuhoff, 2003). Moreover, the registration of certain auditory information in the absence of attention fits with the Bregman (1990) two-stage model of auditory scene analysis, in which unstructured information arriving at the ear is preattentively segregated according to the acoustic properties described above, while controlled processes operate at a later stage of analysis during which individual auditory objects undergo a more detailed and complete analysis. Consequently, the observation that middle latency auditory evoked responses are modulated by the presence of mistuning within a complex sound is consistent with the inclusion of inharmonicity to the list of potential candidates which act as preattentive sound segregation cues. Indeed, in a more general sense, MLAERs may prove to be a useful tool in providing converging evidence for those acoustic properties that fulfill the criteria of preattentive cue and which ultimately assist in the apprehension of multiple objects within an auditory scene.

V. CONCLUSION

The data from experiments 1 and 2 reveal a reliable modulation of the middle latency wave Pa in response to the mistuning of the third harmonic in a complex 200-Hz stimulus. The latency and scalp location of this effect suggests the activity of a neural generator within primary auditory cortex. This finding is consistent with the preattentive registration of inharmonicity and its concomitant use as a cue for concurrent sound segregation. Further work should seek to confirm the spatial origin of the current effect using high-resolution imaging techniques, investigate the interaction between fundamental frequency and the detection of mistuning, and draw further links between early auditory evoked potentials and the later perceptual experience of the individual.

ACKNOWLEDGMENTS

The research reported here was funded by a grant from CIHR and a Premier Research Excellence Award to the second author, and the Baycrest Center for Geriatric Care. The authors wish to thank David Purcell for assistance with the computer simulations and two anonymous reviewers for helpful comments regarding an earlier version of this manuscript.

Alain, C., and Izenberg, A. (**2003**). ''Effects of attentional load on auditory scene analysis,'' J. Cogn Neurosci. **15**, 1063–1073.

Alain, C., Arnott, S. R., and Picton, T. W. (**2001**). ''Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials,'' J. Exp. Psychol. Hum. Percept. Perform. **27**, 1072–1089.

Alain, C., Schuler, B. M., and McDonald, K. L. (**2002**). ''Neural activity associated with distinguishing concurrent auditory objects,'' J. Acoust. Soc. Am. **111**, 990–995.

Amenedo, E., and Díaz, F. (**1998**). ''Effects of aging on middle-latency auditory evoked potentials: A cross-sectional study,'' Biol. Psychiatry **43**, 210–219.

Ballantyne, D. (**1990**). *Handbook of Audiological Techniques* (Butterworth-Heinemann, London).

Beiter, R. C., and Hogan, D. D. (**1973**). ''Effects of variations in stimulus rise-decay time upon the early components of the auditory evoked response,'' Electroencephalogr. Clin. Neurophysiol. **34**, 203–206.

Berg, P., and Scherg, M. (**1994**). ''A multiple source approach to the correction of eye artifacts,'' Electroencephalogr. Clin. Neurophysiol. **90**, 229–241.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sounds* (MIT, London).

Chalikia, M. B., and Bregman, A. S. (**1989**). ''The perceptual segregation of simultaneous auditory signals: Pulse train segregation and vowel segregation,'' Percept. Psychophys. **46**, 487–496.

Eggermont, J. J., and Ponton, C. W. (**2002**). ''The neurophysiology of auditory perception: From single units to evoked potentials,'' Audiol. Neuro-Otol. **7**, 71–99.

Hartmann, W. M. (**1988**). ''Pitch, perception and the segregation and integration of auditory entities,'' in *Auditory Functions: Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 623–645.

Kubovy, M., and Van Valkenburg, D. (**2001**). ''Auditory and visual objects,'' Cognition **80**, 97–126.

Lane, R. H., Kupperman, G. L., and Goldstein, R. (**1971**). ''Early components of the averaged electroencephalic response in relation to rise-decay time and duration of pure tones,'' J. Speech Hear. Res. **14**, 408–415.

Liégois-Chauvel, C., Musolino, A., and Chauvel, P. (**1991**). ''Localization of the primary auditory areas in man,'' Brain **114**, 139–153.

Lin, J. Y., and Hartmann, W. M. (**1998**). ''The pitch of a mistuned harmonic: Evidence for a template model,'' J. Acoust. Soc. Am. **103**, 2608–2617.

Lins, O. G., Picton, T. W., Berg, P., and Scherg, M. (**1993a**). ''Ocular artifacts in EEG and event-related potentials: I. Scalp topography,'' Brain Topogr **6**, 51–63.

Lins, O. G., Picton, T. W., Berg, P., and Scherg, M. (**1993b**). ''Ocular artifacts in recording EEGs and event-related potentials: II. Source dipoles and source components,'' Brain Topogr **6**, 65–78.

McGee, T., Kraus, N., Littman, T., and Nicol, T. (**1992**). ''Contribution of medial geniculate body subdivision to the middle latency response,'' Hear. Res. **61**, 147–152.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1986**). ''Thresholds for hearing mistuned partials as separate tones inharmonic complexes,'' J. Acoust. Soc. Am. **80**, 479–483.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1985**). ''Thresholds for the detection of inharmonicity in complex tones,'' J. Acoust. Soc. Am. **77**, 1861–1867.

Neuhoff, J. G. (**2003**). ''Pitch variation is unnecessary (and sometimes insufficient) for the formation of auditory objects,'' Cognition **87**, 219–224.

Palmer, A. R. (**1990**). ''The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear-nerve fibers,'' J. Acoust. Soc. Am. **88**, 1412–1426.

Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., and Elbert, T. (**1995**). ''Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings,'' Electroencephalogr. Clin. Neurophysiol. **94**, 26–40.

Picton, T. W., and Fitzgerald, P. G. (**1983**). ''A general description of the human auditory evoked potentials,'' in *Bases of Auditory Brain-stem*

Evoked Responses, edited by E. J. Moore (Grune and Stratton, New York), pp. 141–156.
Picton, T. W., van Roon, P., Armilio, M. L., Berg, P., Ille, N., and Scherg, M. (**2000**). ''The correction of ocular artifacts: A topographic perspective,'' Clin. Neurophysiol. **111**, 53–65.
Scheffers, M. T. M. (**1983**). ''Simulation of auditory analysis of pitch: An elaboration on the DWS pitch meter,'' J. Acoust. Soc. Am. **74**, 1716–1725.
Scherg, M., and von Cramon, D. (**1986**). ''Evoked dipole source potentials of the human auditory cortex,'' Electroencephalogr. Clin. Neurophysiol. **65**, 344–360.
Sinex, D. G., Sabes, J. H., and Li, H. (**2002**). ''Responses of inferior colliculus neurons to harmonic and mistuned complex tones,'' Hear. Res. **168**, 150–162.
Skinner, P. H., and Antinoro, F. (**1971**). ''The effects of signal rise time and duration on the early components of the auditory evoked cortical response,'' J. Speech Hear. Res. **14**, 552–558.
Thornton, A. R., Mendel, M. I., and Anderson, C. V. (**1977**). ''Effects of stimulus frequency and intensity on the middle components of the averaged auditory electroencephalic response,'' J. Speech Hear. Res. **20**, 81–94.
Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., and Bloom, F. E. (**1993**). ''Modulation of early sensory processing in human auditory cortex during auditory selective attention,'' Proc. Natl. Acad. Sci. U.S.A. **90**, 8722–8726.
Woods, D. L., and Clayworth, C. C. (**1985**). ''Click spatial position influences middle latency auditory evoked potentials (MAEPs) in humans,'' Electroencephalogr. Clin. Neurophysiol. **60**, 122–129.
Woods, D. L., Alain, C., Covarrubias, D., and Zadiel, O. (**1993**). ''Frequency-related differences in the speed of human auditory processing,'' Hear. Res. **66**, 46–52.
Woods, D. L., Alain, C., Covarrubias, D., and Zadiel, O. (**1995**). ''Middle latency auditory evoked potentials to tones of different frequency,'' Hear. Res. **85**, 69–75.

Cuing effects for informational masking

Virginia M. Richards[a)]
Department of Psychology, 3815 Walnut Street, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Donna L. Neff
Boys Town National Research Hospital, 555 N. 30th Street, Omaha, Nebraska 68124

(Received 21 April 2003; revised 10 October 2003; accepted 20 October 2003)

The detection of a tone added to a random-frequency, multitone masker can be very poor even when the maskers have little energy in the frequency region of the signal. This paper examines the effects of adding a pretrial cue to reduce uncertainty for the masker or the signal. The first two experiments examined the effect of cuing a fixed-frequency signal as the number of masker components and presentation methods were manipulated. Cue effectiveness varied across observers, but could reduce thresholds by as much as 20 dB. Procedural comparisons indicated observers benefited more from having two masker samples to compare, with or without a signal cue, than having a single interval with one masker sample and a signal cue. The third experiment used random-frequency signals and compared no-cue, signal-cue, and masker-cue conditions, and also systematically varied the time interval between cue offset and trial onset. Thresholds with a cued random-frequency signal remained higher than for a cued fixed-frequency signal. For time intervals between the cue and trial of 50 ms or longer, thresholds were approximately the same with a signal or a masker cue and lower than when there was no cue. Without a cue or with a masker cue, analyses of possible decision strategies suggested observers attended to the potential signal frequencies, particularly the highest signal frequency. With a signal cue, observers appeared to attend to the frequency of the subsequent signal. [DOI: 10.1121/1.1631942]

PACS numbers: 44.66.Dc [MRL] Pages: 289–300

I. INTRODUCTION

Under conditions of high stimulus uncertainty, or informational masking (cf. Watson, 1987), detection thresholds can be substantially elevated relative to thresholds measured under conditions of minimal uncertainty. For the detection of a tone added to a masker, stimulus uncertainty can be introduced in a variety of ways. Overall, signal-frequency uncertainty appears to have only modest effects, averaging around 3 dB, for the detection of a tone added to noise and for the detection of an increment in the level of one of many simultaneously presented tones (profile analysis; Green, 1988; Spiegel *et al.*, 1981; Raney *et al.*, 1989). In contrast, uncertainty in the frequency content of standards/maskers can have quite large detrimental effects, both for profile analysis tasks and for the detection of a tone added to a random-frequency, spectrally sparse multitone complex (Neff and Callaghan, 1988; Neff and Dethlefs, 1995; Spiegel *et al.*, 1981; Wright and Saberi, 1999). Presumably, such spectral variation degrades the observers' internal representation of the standard/masker, and therefore the sensitivity to an added signal tone is diminished. Note that variation in the frequency content of the standard/masker is of consequence only if the observer is incorporating energy at nonsignal frequencies. In profile analysis, across-frequency integration is advantageous to the observer, and is in fact the point of the paradigm (Green, 1988). For detection of a tone added to random-frequency, multitone maskers, however, across-frequency integration is nonoptimal in that thresholds would be lower if the observer attended only to the signal frequency.

Presenting a cue prior to a detection trial can reduce uncertainty about stimulus frequencies for the masker or the signal. A preview of an otherwise randomly drawn signal (a signal cue) promotes the allocation of attention at the appropriate frequency (e.g., Hafter *et al.*, 1993; Schlauch and Hafter, 1991). For random-frequency, multitone maskers, we know of no reports testing cue effectiveness directly, although comparisons across studies suggest signal cues can improve performance (Oh *et al.*, 2001). This paper compiles data from two independent laboratories on the effects of signal and masker cues. Across a variety of manipulations and despite some differences in details of stimulus properties and procedures, a consistent picture emerges of effects of masker and signal cues on detection of tones added to random-frequency, multitone maskers.

II. FIXED-FREQUENCY SIGNALS: EFFECT OF A SIGNAL CUE

The data reported here test the effectiveness of a pretrial cue, a preview of the signal, for conditions with a known, fixed-frequency signal at 1000 Hz. The maskers were random-frequency, multitone complexes in which the number of masker components (N) was varied across conditions. Data are combined from independent experiments in two laboratories, one at Boys Town National Research Hospital (BTNRH) and the other at the University of Pennsylvania (UPenn). The conditions tested and methods used were similar, and the results are remarkably consistent. In the first set

[a)]Electronic mail: richards@psych.upenn.edu

of conditions summarized below, results for maskers with two and ten components are from BTNRH; the results for maskers with six components are from UPenn, both using forced-choice, adaptive procedures. In the second set of conditions summarized, the results are from BTNRH, using two- and ten-component maskers with various procedural comparisons to test cue effectiveness. Specifically, masker uncertainty was varied within trials and a single-interval, yes–no procedure was compared to a 2IFC procedure, with and without signal cues. In all these conditions, there was no (physical) uncertainty in the signal frequency other than that presumed to be produced perceptually by pairing the signal with the random-frequency maskers. Because it has been well established that random-frequency maskers with small numbers of components can produce large detrimental effects on detection (e.g., Neff and Dethlefs, 1995; Oh and Lutfi, 1999), it was expected that a pretrial cue to the signal would aid performance. Further, it was expected that the larger the threshold elevation produced by masker-frequency uncertainty, the larger the potential release from masking with a pretrial signal cue.

A. Methods

1. Observers

Eight (BTNRH) or ten (UPenn) normal-hearing observers, ranging in age from 19–29 years, participated. All had thresholds in quiet of 15 dB HL or better for octave frequencies from 250 to 8000 Hz, and were paid for their participation. Observers were tested in individual sound-attenuated rooms in groups of up to four (BTNRH) or individually (UPenn). Stimuli were presented diotically over Sennheiser HD250-II (BTNRH) or HD410SL (UPenn) headphones.

2. Stimuli

The signal was always a 1000-Hz sinusoid. The maskers were multitone complexes with two, ten (BTNRH), or six (UPenn) components. Masker components were sinusoids drawn at random from a uniform frequency distribution on a linear frequency scale from 300–3000 Hz (BTNRH) or a logarithmic frequency scale ranging from 100–6000 Hz (UPenn), excluding a ''protected'' region of $\pm 8\%$ around the signal frequency to reduce potential contributions of energy-based masking. Component phases were drawn at random and component amplitudes were equated. Masker level (total power) was similar across labs, at 60 dB SPL for two- and ten-component maskers (BNTRH) and 58 dB SPL for the six-component maskers (UPenn). Total stimulus durations were 200 ms (BTNRH) or 102 ms (UPenn), both with 5-ms, cosine-squared, onset–offset ramps. In both labs, stimuli were digitally generated and presented through two channels of a 16-bit DAC using a sampling rate of 20 000 Hz, and low-pass filtered at 4000 Hz (BTNRH) or 7000 Hz (UPenn) using matched filters (KEMO VBF 8 or Stewart VBF 10M). The signal, when presented, was synchronously gated with the masker. The signal cue, when presented, was exactly the signal to be used in the upcoming trial, that is, a 1000-Hz tone at the level determined by an adaptive procedure. The interval between the cue and the first listening interval as well as the interstimulus intervals were 500 ms (BNTRH) or 600 ms (UPenn).

3. Procedures and conditions

Despite a number of differences in details, the procedures of the two labs were similar overall. An adaptive procedure was used that estimated the 71% point on the psychometric function, that is, signal level was altered using a 2-down, 1-up algorithm (Levitt, 1971). Observers indicated the presence of a signal with a keyboard press, and got visual correct-answer feedback. The initial signal level was approximately 10 dB above the ultimate threshold. After three reversals the 4-dB step size was reduced to 2 dB. The track was terminated after 80 (BTNRH) or 50 (UPenn) trials, and the threshold determined from the average of all reversals (BTNRH) or the last even number of reversals (UPenn) after the first three. At BTNRH, 20 replications of each condition were completed, and the first two discarded as practice. Thus, thresholds reported represent averages of 18 blocks of 80 trials. At UPenn, the threshold estimates reported vary due to observer availability, but are based on averages of at least 20 but typically 40–45 blocks of 50 trials. At BTNRH, data were also collected for two- and ten-component maskers, with and without a signal cue, using three procedures: two-interval, forced-choice (2IFC) with the masker randomly drawn for each of the intervals (FC-R), 2IFC with the masker drawn at random for each new trial, but repeated for the two intervals within a trial (FC-F), and single-interval, yes–no, with maskers randomized each trial (YN). The signal was as likely to be presented in either interval when the 2IFC procedure was tested, and the signal was as likely as not to be present for the YN procedure. These conditions were tested in random order using a Latin-square design. At UPenn, data were collected for the six-component masker conditions, with and without a signal cue, as part of a larger set of six conditions described in more detail in a later section. The UPenn condition matched the FC-R condition of BTNRH (2IFC with maskers randomized with every presentation).

B. Results and discussion

1. Vary number of masker components

Figure 1 shows the threshold estimates in dB SPL for the FC-R condition, with masker frequency randomized on each presentation. The filled squares are for the conditions in which a signal cue preceded each trial and the unfilled squares are for conditions without a cue. The three panels show data for maskers with two, six, and ten components, respectively. For each panel, data for individual observers are rank ordered from highest to lowest masked thresholds without the cue. The average threshold across all individuals is plotted to the far right of each panel. Error bars are standard errors of the mean, across threshold replicates for individuals or across observers for the average. The dashed horizontal line provides an estimate of the signal level required for overall stimulus loudness to support the detection of the signal; that is, adding a signal with a level at or above the line

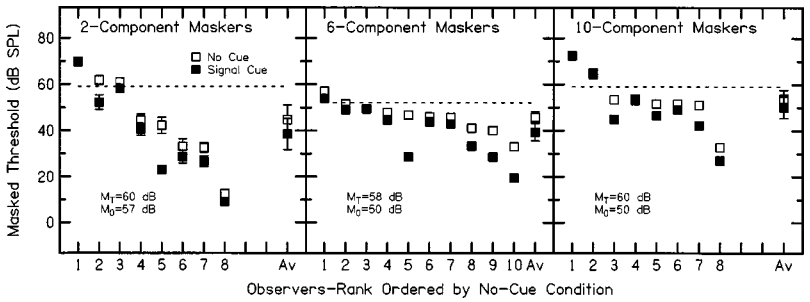


FIG. 1. Thresholds are plotted for individual observers rank ordered by thresholds in the no-cue condition, and for the average across observers (Av). Panels are for two, six, and ten-component maskers, left to right. Filled symbols show thresholds when a signal cue precedes the trial and unfilled symbols indicate thresholds estimated without a cue. Error bars indicate the standard error of the mean across replicate threshold estimates (individual observers) or across observers (averaged results). The dashed horizontal line is an estimate of loudness-based thresholds. M_T and M_0 indicate the total masker level and the level of the individual masker components, respectively.

yields an increment of a least 1 dB relative to the level of the masker alone. This line is lower for the six-component masker because the masker level is lower. Both overall masker level (M_T) and component level (M_0) are given in each panel.

Overall, the results were consistent with other studies of multitone maskers with random-frequency maskers and a fixed-frequency signal (e.g., Neff and Dethlefs, 1995) in that there were substantial differences across observers in performance, particularly for the two-component masker condition. The addition of a cue resulted either in no change in threshold or in a range of threshold reductions with a maximum of 18 dB. Across all observers and maskers, the average threshold reduction with a signal cue was 5.4 dB. For maskers with two, six, and ten components, the pretrial cue resulted in threshold reductions of 6.2, 6.5, and 3.7 dB, respectively. A two-tailed, matched-sample *t*-test indicated that each of these differences was statistically significant ($t=3.0$, 3.4, and 2.6; $p<0.03$ for Ns of 2, 6, and 10, respectively).[1] As a comparison, when signal frequency was uncertain (randomized) for tonal signals masked by Gaussian noise or in profile-analysis tasks, thresholds were elevated by about 3 dB (Green, 1961; Spiegel *et al.*, 1981; Raney *et al.*, 1989). This suggests that randomizing the frequencies of these multitone maskers with a fixed-frequency signal had a greater detrimental effect on detection than randomizing the signal frequency itself with broad-spectrum maskers.

The rank ordering of observers makes it clear that, with or without a signal cue, only a few of the poorest observers had thresholds at or above the ''loudness limit.'' Thus, most observers did not appear to be using loudness as a basis for signal detection, and the few who may have done so showed no benefit of a signal cue (e.g., $L1$ for $N=2$ and $L1$ and $L2$ for $N=10$). Similarly, one would expect a ''floor'' in the size of cuing effects for observers with the best performance, whose thresholds reflected little detrimental effects of the masker-frequency uncertainty (e.g., $L8$ for $N=2$). Most observers in this study, however, had performance between these extremes, and their results suggested that performance without a cue was not a strong predictor of how much they would benefit from a cue. In Fig. 1, there are a number of examples of observers with essentially identical performance in no-cue conditions whose cued thresholds differed (e.g., $L4$ vs $L5$ for either $N=2$ or $N=6$). Correlations of the amount of masking release (thresholds without a cue minus thresholds with a cue) with thresholds without a cue for $N=2$, 6, and 10 were 0.1, 0.4, and 0.5, respectively, none of which was statistically significant ($p>0.05$). Overall, these data suggest that the use of a pretrial cue to the signal does not degrade performance, and improves performance for the majority of observers, quite substantially in some cases.

2. Vary procedures and degree of masker randomization

Figure 2 plots threshold estimates collected using three different procedures. For each procedure, a pretrial signal cue was either present or not. The procedures differed in the stimuli available to the subjects for comparison across or within trials. In the two 2IFC procedures, new masker samples were randomly drawn for both intervals within a trial (FC-R) or fixed within a trial and randomly drawn across trials (FC-F). In the single-interval, yes–no (YN) procedure, one masker sample was randomly drawn for each

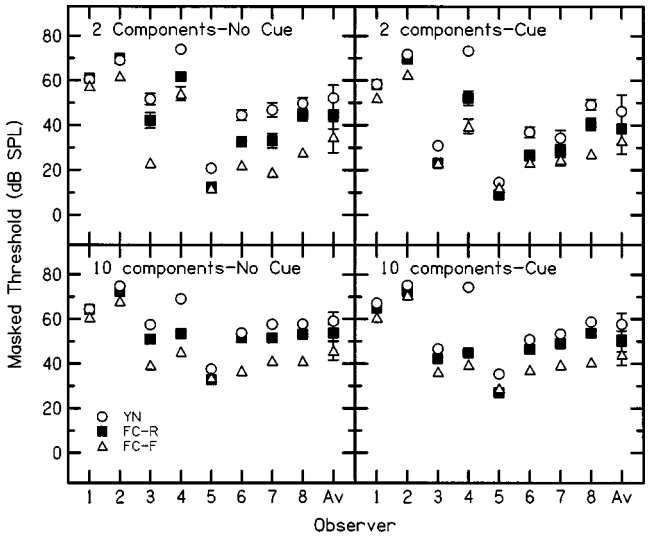


FIG. 2. Thresholds are plotted for the individual observers and averaged across observers (Av). Panels are for two-component (top row) and ten-component (bottom row) maskers. The left panels show thresholds without a pretrial signal cue and the right panels plot thresholds when a pretrial signal cue precedes the trial. The different conditions, yes/no (YN), forced-choice random (FC-R), and force-choice fixed (FC-F) are indicated using circles, filled squares, and triangles, respectively. Error bars are as in Fig. 1.

trial. Each panel in Fig. 2 presents results for eight observers and the average across observers, with observers ordered as tested (not rank ordered) for ease of comparing observers' performance across panels. Data for the FC-R condition with and without a cue were replotted from Fig. 1 as filled squares in all panels, with unfilled circles and triangles for the YN and FC-F conditions, respectively. With or without a signal cue, masked thresholds were highest for the YN conditions, and lowest for the FC-F conditions. This suggests that observers benefit from comparisons of masker samples across intervals to aid signal detection even when the maskers are randomly sampled across two listening intervals. It should be noted that the two methods, YN and FC, theoretically ''track'' different values of d'. For YN trials the d' value is expected to be somewhat higher[2] than 1, whereas for the FC trials the d' value is expected to be near 0.8. Even so, the impact of the cue within procedures (FC-R and YN) and comparisons of FC-R and FC-F conditions suggest a benefit from comparisons of masker samples across intervals. These comparisons are considered in more detail below.

A comparison of the range of thresholds for two vs ten components indicates large differences in performance across observers for both sets of data, but greater scatter for the two-component maskers with or without the signal cue. This is consistent with past studies that used similar stimuli and procedures but no cue (e.g., Neff and Dethlefs, 1995; Oh and Lutfi, 1999). Also consistent with earlier studies, the ten-component masker consistently produced more masking than the two-component masker. Averaged across all procedures, ten-component maskers produced 10.2 dB more masking, with differences ranging from 7.0 dB for YN-no cue to 11.6 for FC-R-cued. Thus, the addition of a signal cue did not appear to change the typical form of the masking function observed when the number of components is manipulated (Neff and Green, 1987; Neff and Dethlefs, 1995; Oh and Lutfi, 1999). This study, however, did not include conditions that would reveal the nonmonotonic function (i.e., the reduction in masking that ultimately results as larger Ns are tested). A repeated-measures ANOVA indicated significant effects of presentation method [$F(2,14)=20.6$; $p<0.01$], number of components [$F(1,7)=13.4$; $p<0.01$], and signal cue [$F(1,7)=7.4$; $p<0.05$]. Among the available interactions, only the interaction between number of components and the presence of a cue was statistically significant [$F(1,7)=6.9$; $p<0.05$], reflecting a larger effect of the cue when $N=2$ than when $N=10$. Although significant, the effect was modest, in that for $N=2$ adding a cue reduced the average thresholds 4.6 dB compared to 2.2 dB for $N=10$.

The release from masking produced by the signal cue is more easily assessed in Fig. 3, in which the difference in thresholds with vs without a cue is plotted as a function of the masked threshold without a cue. The format (left column for $N=2$, right column for $N=10$) and the symbols for the three procedures remain the same as in Fig. 2. Within each panel, the eight identical symbols per procedure represent the eight observers. Points falling at or below the horizontal dashed lines indicate conditions where the cue either had no effect on thresholds or degraded performance. Table I provides a summary of the range of cuing effects and the mean

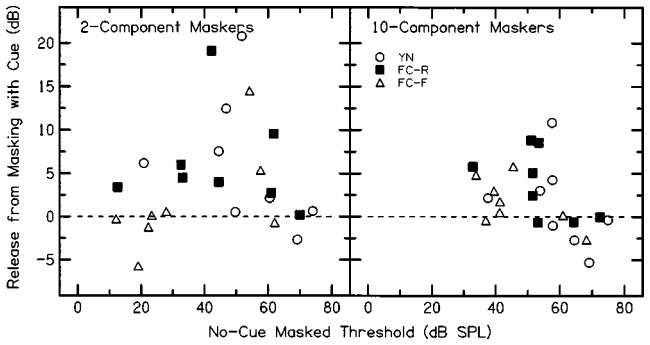


FIG. 3. Cued-based release from masking is plotted as a function of thresholds obtained in the absence of cues. The left panel is for then two-component maskers and the right panel is for the ten-component maskers. Circles, filled squares, and triangles are for the YN, FC-R, and FC-F conditions. Multiple symbols indicate the amount of masking for individual observers.

observed across observers for each of the three procedures and the two types of maskers. As expected from past studies, thresholds were more regular across observers when the masker was comprised of ten rather than two components (e.g., Neff and Dethlefs, 1995). In addition, there was a smaller impact of a signal cue on thresholds for the ten- than for the two-component masker. Both effects are clearly seen in Fig. 3. No-cue thresholds were restricted to a higher range for $N=10$ (30 dB or greater), and both the maximum and average release from masking was smaller for $N=10$ than for $N=2$, except for the FC-F condition, as shown in Table I. For all conditions, there were observers for whom the cue either had no effect or even degraded performance up to 6 dB. In many conditions, however, particularly the FC-R and YN conditions with maskers randomized with each presentation, the cue provided significant improvements in performance. Unfortunately, which observers will benefit from a signal cue is not easily predicted. As illustrated by the leftmost and rightmost points in each panel in Fig. 3, observers

TABLE I. A summary of the release from masking (in dB) associated with the addition of a signal cue is presented for each of the three procedures (YN, FC-R, FC-F) and maskers with two and ten components. For each procedure, the minimum, maximum, and mean release from masking across the eight observers is given. Numbers in parentheses (SE) are standard errors of the means.

	Masker Components	
Procedure	2	10
YN		
Minimum	−2.7	−5.3
Maximum	20.8	10.8
Mean (SE)	6.0(2.7)	1.4(1.8)
FC-R		
Minimum	0.2	−0.7
Maximum	19.1	8.8
Mean (SE)	6.2(2.1)	3.7(1.4)
FC-F		
Minimum	−5.7	−2.7
Maximum	14.5	5.8
Mean (SE)	1.6(2.1)	1.6(1.0)
Grand mean (SE)	4.6(1.5)	2.2(0.7)

at either extreme of the no-cue distribution of thresholds showed the least benefit of the cue. This was not surprising for the best performers, as a floor effect limits the range of cuing effects. However, observers who had the most difficulty with the task also did not benefit from the signal cue. The fact that a cue to signal frequency and level did not aid these observers again suggests that these observers might have been using loudness as the basis for their detection decisions (see Fig. 1) and failed to adopt a more efficient strategy when a cue is added. Except for these extreme performers, however, the signal cue had a large effect for the majority of observers, improving thresholds up to 20 dB for individuals.

For the FC-F conditions, no-cue thresholds were lower than for the corresponding FC-R conditions. This is consistent with previous studies indicating that observers can take advantage of comparing masker samples across intervals to improve performance with these multitonal maskers (e.g., Neff and Callaghan, 1988; Neff and Dethlefs, 1995). For observers who did not show much improvement for FC-R compared to FC-F, adding a signal cue improved performance by up to 14 dB for FC-F. For the FC-F procedures, the cue produces more consistent releases from masking.

The results indicate that a pretrial signal cue can improve thresholds for at least two types of multitone maskers and several presentation methods. Note that across the conditions and observers tested, thresholds measured with and without a cue were highly correlated ($r=0.95$), that is, the results showed consistent patterns for masker type and procedure. In contrast, the correlation between no-cue thresholds and the release from masking associated with the signal cue was small ($r=-0.12$). Even excluding the most sensitive and insensitive observers, the relationship was weak. It appears that a pretrial signal cue provided a release from masking that was more or less independent of task and/or amount of masking present in the base condition. This result is in accord with other studies in which stimulus properties were altered in an effort to reduce the amount of informational masking (Neff and Dethlefs, 1995; Neff, 1991). For example, when the signal onset is delayed relative to the masker onset, there is substantial reduction in masking, and a corresponding strong correlation between the large amount of masking in conditions without signal onset delay and the release in masking associated with the delay. This reflects the relatively low thresholds measured for *all* observers when the signal onset is delayed relative to the masker onset. For other stimulus manipulations that produced more modest or no reduction in informational masking, (as observed here), the amount of release from informational masking was not clearly correlated with the base threshold (Neff and Dethlefs, 1995; Neff, 1991).

On average, across all observers and conditions, the cue effect was modest, with a reduction in masking of only 3.4 dB. For individuals, however, the cue effects can be very large, suggesting that if best performance is desired, a pretrial cue to the signal is worthwhile.

III. RANDOM-FREQUENCY SIGNALS: EFFECT OF A SIGNAL CUE AND LEVEL VARIATION

The results presented above for conditions with a fixed-frequency signal indicated that a signal cue can lower detection thresholds with random-frequency, multitonal maskers. In this section, the effect of a pretrial signal cue was examined for conditions in which the signal frequency was randomly chosen on a trial-by-trial basis. For the detection of a tone added to Gaussian noise and for profile analysis tasks, signal frequency uncertainty has little impact on thresholds (Green, 1961; Spiegel *et al.*, 1981; Raney *et al.*, 1989). With the random-frequency, multitonal maskers, however, randomizing the signal frequency would be expected to dramatically increase the difficulty of the task because a pretrial signal cue would provide the only predictor of signal frequency. In addition, conditions were tested in which the overall levels of the stimuli were randomly chosen prior to each stimulus presentation. If observers incorporate changes in overall loudness, then randomizing overall stimulus levels prior to each stimulus presentation will degrade performance.

A. Methods and procedures

The results from UPenn reported above in Sec. I for six-component maskers and fixed-frequency signals are part of a larger data set. This section reports results for the six-component maskers in six conditions. Four conditions compared performance for fixed- vs random-frequency signals with and without a signal cue. When random, the frequency of the signal was chosen from the same range of frequencies as the masker components (i.e., a uniform distribution on a logarithmic frequency scale ranging from 100 to 6000 Hz).[3] These conditions are labeled as follows: *Fix* (fixed signal frequency with no cue), *FixCue* (fixed signal frequency with cue), *Ran* (random signal frequency with no cue), and *RanCue* (random signal frequency with cue). The *Fix* and *FixCue* conditions were also presented in Sec. I. In two additional conditions, thresholds were obtained without a cue (*Fix* and *Ran*) in the presence of level randomization. These conditions are labeled as *LevFix* (random stimulus levels, fixed signal frequency) and *LevRan* (random stimulus levels, random signal frequency).

The methods, procedures, and observers were as described earlier for the UPenn site. Briefly, ten observers were tested with the FC-R procedure with six-component multitone maskers. Observer 2 did not complete the conditions with level randomization. For level randomization, the stimuli were amplified or attenuated using random draws from a uniform distribution ranging from -15 to 15 dB. In the staircase procedure, stimulus levels greater than 90 dB SPL were not allowed. If the staircase procedure continued to reach this upper limit for an observer after at least 250 trials, the condition was not completed for that observer. Because the *LevRan* condition proved very difficult, this happened frequently.

B. Results and discussion

Figure 4 compares thresholds (dB SPL) with and without a signal cue for fixed- and random-frequency signals

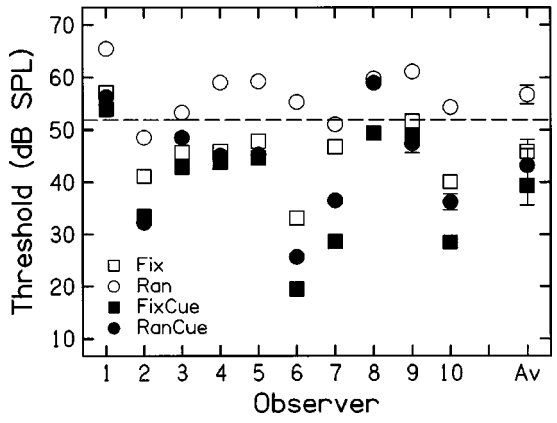


FIG. 4. Thresholds are plotted for the individual observers and averaged across observers (Av). Filled symbols show thresholds estimated when a signal cue preceded the trial and unfilled symbols show thresholds without a cue. The signal frequency is either fixed at 1000 Hz (squares) or random (circles). Error bars indicate the standard error of the mean across replicate threshold estimates (individual observers) or across observers (averaged results).

with no level variation. Results for ten observers and the average across observers (far right) are plotted. Data for *Fix* and *FixCue* conditions are the same as in Fig. 1 (open and filled squares for no cue and cue conditions in both figures), but the observers are not rank ordered to aid comparisons across conditions for individuals. Error bars are standard errors of the mean across threshold replicates (individual observers) or across observers (average). Data for the *Ran* and *RanCue* conditions are indicated using open and filled circles, respectively. As in Fig. 1, the dashed horizontal line provides an estimate of the signal level required for overall stimulus loudness to support the detection of the signal; that is, the addition of a signal with a level greater than 52 dB SPL yields an increment of a least 1 dB relative to the level of the masker alone.

As is clear from Fig. 4, Observer 1 was unique in her high thresholds for all four conditions. When the signal frequency was random and there was no cue (*Ran*, open circles), thresholds for nearly all of the observers were above the ''loudness limit.'' This is consistent with expectations based on the results of Neff and Green (1987). They found that observers could not reliably discriminate between six- vs eight-component, equal-amplitude complexes when the component frequencies were randomly chosen and the stimulus levels were consistent across intervals. The current results demonstrate that observers could not discriminate between stimuli composed of six vs seven random tones, even when the ''extra'' signal component's amplitude was different from that of the other (masker) components. Compared to the Neff and Green study, stimulus levels were higher when the signal (extra) tone was added.

As noted in Sec. I, when signal frequency was fixed, adding a signal cue led to an average reduction in threshold of 6.5 dB. By comparison, when signal frequency was random, the average threshold reduction with the cue was larger at 13.5 dB. Note, however, that this larger cuing effect may reflect a change in strategies (e.g., a reduction in emphasis on changes in loudness) and not an improvement in a single strategy. Regardless, there is no question that the cue was highly beneficial for this difficult task. When a signal cue was present to aid observers, randomizing the signal frequency increased thresholds an average of 3.9 dB compared to when the signal frequency was fixed.

TABLE II. A summary of masked thresholds (dB SPL) for eight observers (Observers 3–10 in Fig. 3 from UPenn) and the average across observers for conditions with and without level variation. The masker had six random-frequency components and no signal cue was used. Standard errors are indicated in parentheses. These conditions test the degree to which observers rely on overall loudness for detection of the fixed-frequency, 1000-Hz signal.

Observer	No level variation	Level variation
3	45.7 (0.8)	42.3 (1.2)
4	45.9 (0.7)	48.0 (0.6)
5	47.8 (0.6)	48.2 (0.6)
6	33.1 (1.1)	33.9 (0.7)
7	46.7 (0.7)	48.2 (0.6)
8	49.4 (1.1)	48.8 (0.8)
9	51.6 (1.4)	58.4 (1.8)
10	40.0 (1.2)	35.0 (1.5)
Average	45.0 (2.1)	45.4 (2.9)

One concern was whether the signal cue would introduce a strong response bias, particularly for the difficult random-signal conditions. For example, did observers more frequently indicate the presence of a signal in the first interval when a signal cue was present? Responses for Observers 4–10 were examined, and the results revealed no reliable differences in the proportion of times they indicated interval 1 as the signal interval dependent on whether or not the FC-R trial was preceded by a cue.

The effect of adding overall level randomization for conditions with a fixed-frequency signal (*FixLev*) is summarized in Table II, which compares thresholds with and without level randomization for Observers 3–10 (numbered as in Fig. 4) and the average. Data were omitted for Observer 1 because she had signal levels that frequently exceeded the 90-dB upper limit for the staircase procedure. Observer 2 did not run this condition. As shown in Table II, thresholds were essentially the same whether or not the levels were randomized, and a matched *t*-test failed to indicate a significant difference.

No data are presented for the *RanLev* condition because the task proved so difficult that seven of the nine observers could not attain thresholds below the 90-dB staircase limit. For the two observers from whom thresholds were estimated, the thresholds were high (68.5 and 67.5 dB, respectively).

In summary, when a fixed-frequency signal was added to random, multitone maskers, observers appeared not to rely on overall loudness to detect the added tone. This was reflected in two ways: (1) thresholds were low compared to the 1-dB loudness limit; and (2) level randomization failed to alter thresholds compared to conditions without level randomization. A signal cue led to lower thresholds compared to when there was no cue. It seems likely that this indicated an improvement in strategy for cued conditions, rather than a substantial change in detection strategy. When the signal frequency was random and the overall levels were randomized, thresholds were either very high or not measurable. In this condition, a pretrial signal cue identifying the signal frequency led to substantial reductions in thresholds, possibly

reflecting a reduction in reliance on loudness for detecting the added signal.

IV. RANDOM-FREQUENCY SIGNALS: EFFECT OF A SIGNAL OR MASKER CUE AND VARYING INTERSTIMULUS INTERVAL

In these conditions the interstimulus interval (ISI), the time between the cue offset and the trial onset, was systematically varied for random-frequency, multitone maskers combined with a random-frequency signal. Performance was assessed without a cue, with a pretrial cue for the signal, and with a pretrial cue for the masker. The primary goal of this experiment was to determine whether a pretrial cue might reduce thresholds by encouraging the formation of an auditory stream, which would help distinguish masker from signal. When a signal cue is used, an auditory stream might be formed when the subsequent stimulus has an added signal, but not when the subsequent stimulus lacks the signal. When a masker cue is used, the opposite may occur (i.e., the signal segregates because the masker repeats). Auditory streaming, then, may allow observers to more easily detect an added tone when a pretrial cue, either for the signal or the masker, precedes a trial.

Previous work by Kidd and colleagues (Kidd *et al.*, 1994, 2002a) has shown that perceptual grouping may aid signal detection in conditions of high uncertainty. They used a multiburst paradigm in which the stimuli were several multitone maskers (or signal-plus-maskers) played in rapid succession within each interval. Thresholds were lower when the multiple bursts were composed of randomly drawn multitone maskers than when the same masker was repeated across bursts. They attributed the advantage to the formation of a signal stream; that is, the signal component, which had a constant frequency across bursts, perceptually segregated from the random-frequency masker components. Recently, Kidd *et al.* (2002b) have shown that the closer in time the bursts, the stronger the sense of signal coherence and the lower the threshold. For example, for eight bursts, reducing the time between the bursts from 400 to 0 ms produced, on average, a 25-dB reduction in thresholds. In this experiment, we examine whether perceptual segregation might be a factor in cuing effects by manipulating the time between the signal or masker cue and the onset of the trial.

A. Methods and procedures

To make the task more similar to that used by Kidd *et al.* (2002b), there were several changes in the procedures compared to the other UPenn experiments (Sec. I). The dominant changes were that a YN procedure was adopted and the stimuli were reduced to 51 ms in duration. Masker components were drawn from a range of 200–5000 Hz on a logarithmic scale and the signal to be detected was drawn at random from five potential values; 275, 525, 995, 1895, and 3630 Hz. In the adaptive staircase procedure, independent tracks for these five signal frequencies were interleaved such that in 250 trials there were 50 trials for each signal frequency. The initial 6-dB step size was reduced to 3 dB after the third reversal.

There were three conditions: no cue, signal cue, and masker cue. In the signal- and masker-cue conditions, the cue was an exact preview of the subsequent signal (when present) or masker, respectively. For the signal-cue conditions, ISIs of 5, 50, 100, 250, and 500 ms were tested. For the masker-cue conditions, ISIs of 5, 100, and 250 ms were tested.

Six UPenn observers, aged 19–30 years, participated. All had thresholds in quiet of 15 dB HL or better for octave frequencies ranging from 250 to 8000 Hz. Observers were initially trained to detect each of the five potential signal tones in quiet. Data collection was then blocked by condition with half of the observers beginning with the signal-cue condition and the other half beginning with the masker-cue condition. Within the masker- and signal-cue conditions, data collection for the different ISIs and for the no-cue condition were blocked, but tested in unique order across observers. Twenty blocks of trials were run for each condition, yielding 20 separate tracks for each of the five signal frequencies. Of these, threshold estimates from the last 15 tracks were averaged to provide a final threshold estimate.

An adaptive algorithm tracked a constant percent correct, which for a YN task is not assured to lead to constant levels of d'. Therefore, after the experiment was finished, psychometric functions were reconstructed from the adaptive tracks. Each trial was specified as being associated with a particular signal frequency and signal level even when there was no signal. As a result, false-alarm rates were based on trials associated with the particular signal frequency and level.[4] Hit and false-alarm rates were converted to z scores, and d' values estimated as the difference between hit and false-alarm z scores. The data were further reduced by removing signal levels yielding d' values less than 0 or greater than 3 and signal levels for which there were fewer than 20 trials. The psychometric functions were fitted using a linear fit that minimized the weighted squared error between the predicted and observed d's, where the weights were the inverse of the number of trials contributing to the estimate (Dai, 1995; Press *et al.*, 1986, page 504). The result is a line relating d' and signal level in dB SPL. This approach provided an alternative estimate of threshold, which corresponded to the signal level at which $d'=1$.

B. Results and discussion

1. Alternate threshold and bias estimates

The first question addressed was how the thresholds from the adaptive procedure (which estimated 71% accumulated hit and correct rejection rates) compared to the thresholds derived from the psychometric functions (which estimated signal level for $d'=1$).[2] The results indicated that thresholds for the two procedures were stable relative to one another. In the no-cue condition, the d'-based threshold was consistently 2–3 dB lower than the track-based threshold. In the masker-cue condition, the difference was stable at approximately 4 dB. In the signal-cue condition, however, the linear fits for many of the observers were consistently poor for the 5-ms ISI conditions and occasionally poor for the 50-ms ISI conditions, because the psychometric functions

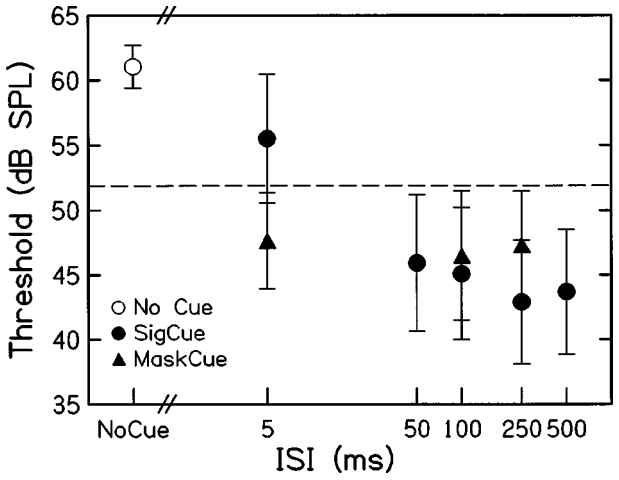


FIG. 5. Average thresholds are shown for the no-cue condition (unfilled circle), the signal-cue condition (filled circles), and the masker-cue condition (filled triangles). For the latter two conditions, thresholds are plotted as a function of ISI. Error bars show the standard error of the mean across observers (6 observers). The dashed line shows the signal level required to increase the masker stimulus by 1 dB.

did not monotonically increase with signal level. Overall, the staircase procedure was judged to provide at least as good ''threshold'' estimates across all conditions as the fitted psychometric functions; thus, the thresholds reported below are based on the original staircase data.

Again, as part of the data analysis, the possibility of response biases was examined. It might be the case that observers' tended to respond ''no signal'' more or less often depending on the type of cue that preceded the trial. Across observers, the proportion of ''no signal'' responses was 0.54. On average, the proportion of ''no-signal'' responses was largest in the masker-cue condition and essentially the same in both the signal-cue and no-cue conditions (0.57, 0.53, and 0.52, respectively). An ANOVA indicated no significant main effect of cue type on bias [$F(2,10)=3.6$, $p>0.06$].

2. No-cue condition

Data in the no-cue condition were collected in both the context of the signal-cue conditions and the context of the masker-cue conditions. Because the no-cue threshold estimates did not depend on the context or the order in which the conditions were run, all estimates for the no-cue conditions were combined. The values presented here represent the average of 30 threshold estimates. Further, an ANOVA failed to reveal a main effect of the five signal frequencies, so thresholds were also averaged across signal frequency. In Fig. 5, masked thresholds are summarized for the three cue conditions (no cue-open circle; signal cue-filled circles; masker cue-filled triangles) as a function of the ISI. Data were averaged first across frequencies and then across observers; thus, error bars indicate the standard errors of the mean across observers. As in earlier figures, the dashed line indicates the ''loudness limit,'' that is, the signal level required to increase the masker level by 1 dB. Masked thresholds with the random signal frequencies without a cue were approximately 4 dB higher than the thresholds estimated using random signal frequencies in the previous experiment (see Fig. 4). This increase was not unexpected, however, given both the shorter stimulus duration used here (51 vs 102 ms), and the increase in thresholds observed in earlier conditions for YN compared to 2IFC procedures (see Fig. 2). The average no-cue threshold was well above the loudness limit, suggesting that loudness cues may have contributed to the detection strategy.

3. Signal- and masker-cue conditions

The filled symbols in Fig. 5 show the average thresholds for the signal-cue (circles) and the masker-cue (triangle) conditions as a function of the ISI. When a signal cue preceded the YN trial, thresholds were higher for the 5-ms ISI than for any of the longer ISIs tested. For ISIs longer than 5 ms, no effect of ISI was apparent for either these average results or for any of the individual observer's results. The large error bars in Fig. 5 reflect large individual differences in the size of the cuing effect. On average, the signal cue reduced thresholds by 14.4 dB relative to no-cue thresholds. For ISIs larger than 5 ms, the average cuing effect was 16.6 dB. The smallest cuing effects (for Observers 5 and 6) were 3–4 dB. For the other four observers the cuing effects were large, ranging from 11 to 30 dB. Recall that when a signal cue preceded the trial, many of the psychometric functions estimated for the 5-ms ISI condition were not monotonically increasing with level. This suggests that the track-based thresholds were also unreliable. Both the flat psychometric functions and the small cuing effect for this 5-ms condition might be attributable to the interplay of conflicting effects of cuing and some form of forward masking (Moore and Glasberg, 1983; Moore, 1981; Zwicker, 1984; Neff, 1991). If the signal cue partially masked the signal or could be perceptually confused with the signal, increases in signal level would not necessarily lead to better performance because the signal and cue were always presented at the same levels.

When the pretrial cue was a preview of the masker, on average the masker cue reduced thresholds by 13.9 dB. This is similar to the overall average of 14.4 dB for the signal cue. For the masker cue, neither the individual data nor the averaged data indicated an effect of ISI. Unlike the results with a signal cue, the masker cue produced a large reduction in masking even for the shortest (5-ms) ISI. This provides further support for the possible negative influence of partial masking for the 5-ms signal-cue condition. Again, there were large individual differences in the effectiveness of the masker cue, as indicated by the large error bars in Fig. 5. For one observer (Observer 6), there was no impact of a masker cue. For the other five observers, however, the effect for masker cues was similar to that for signal cues, ranging from 12 to 26 dB.

A repeated-measures ANOVA was applied to the results for conditions tested with both signal and masker cues (e.g., ISIs of 5, 100, and 250 ms). Cue type, ISI, and signal frequency were fixed variables. There was a significant main effect of signal frequency [$F(4,20)=12.6$; $p<0.001$]. On average, the threshold for the 3630-Hz signal was 8 dB lower than the thresholds for the other four frequencies, a frequency effect that was not present in the no-cue condition. As shown in Table III, which summarizes thresholds as a function of signal frequency, thresholds were lower for the

TABLE III. Thresholds averaged across ISIs and then across observers are shown for each of five signal frequencies. Standard errors of the mean are shown in parenthesis (across observers, after averaging across ISIs).

	Signal frequency (Hz)				
	275	525	995	1895	3630
Signal cue	48.1 (4.2)	47.3 (4.6)	48.3 (4.7)	49.4 (4.8)	40.1 (4.4)
Masker cue	50.5 (3.6)	48.5 (4.4)	47.6 (5.1)	47.8 (4.3)	41.3 (3.3)

3630-Hz signal for both signal and masker cues. The reason for this is not clear. It is unlikely that this effect reflects changes in the audibility of either the masker or signal cue as a function of frequency. The audibility of either cue is not in question, even if the signal itself is partially masked by the cue. Because 5000 Hz was the upper boundary of the range of potential masker frequencies, perhaps this is an ''edge'' effect, in that listeners could more easily identify the signal as higher than many of the masker components. Another possibility is that when the highest signal frequency was used, observers detected an upward shift in pitch for the signal-plus-masker compared to the masker alone. The fact that no effect of signal frequency was observed for the no-cue conditions seems to indicate some sort of relative frequency comparison across cue and trial intervals.

Not surprisingly, given the higher thresholds for the 5-ms ISI for the signal-cue conditions, both the main effect for ISI and the interaction of ISI and cue type were highly significant [$F(2,10)=7.5$; $p<0.01$, and $F(2,10)=13.11$, $p<0.005$, respectively]. Finally, the main effect of cue type was *not* significant. Overall, these data did not support our initial expectation that the cues might lead to lower thresholds due to streaming. Both cue types aided observers, but ISI had little effect, except to elevate thresholds for the shortest ISI for the signal cue.

To summarize the results of this experiment, five of the six observers showed improved performance with signal and masker cues under conditions of joint signal- and masker-frequency uncertainty. This was consistent with the results presented earlier for the effect of a signal cue with fixed- and random-frequency signals. For the shortest (5-ms) ISI tested, the signal cue was less effective in reducing thresholds than the masker cue, possibly due to some form of ''forward'' masking. For the remaining ISIs, cuing effects were the about same regardless of cue type. These results differ from the pattern of results obtained by Kidd *et al.* (2002b) with random-frequency, multicomponent maskers in multiple-burst paradigms, in which the closer together in time the bursts are played, the lower the threshold. The current results, using single added ''bursts'' of masker or signal as a cue, showed either no impact of ISI or a negative effect for the signal cue and the shortest ISI. There are clearly substantial and important differences between the multiple-burst and the YN-cuing procedures. However, the difference in results suggests that fundamentally different mechanisms underlie improved detection based on cues vs improvements based on multiple repeating signals or maskers. If the advantage of the multiple-burst paradigm is due to the perception of stimulus coherence and auditory streaming (Kidd *et al.*, 2002a), it seems the advantage of masker and signal cues in the YN procedure should *not* also be attributed to the formation of auditory streams.

C. Analysis of detection strategies

Post hoc trial-by-trial analyses of the data were used to discern possible detection strategies used by observers with signal and masker cues. It has been argued that a signal cue can direct an observer's attention to a cued frequency (e.g., Penner, 1972; Dai *et al.*, 1991; Schlauch and Hafter, 1991) and additionally, that the cue need not be veridical relative to the ultimate signal frequency (e.g., Hafter *et al.*, 1993). The present analysis evaluated the strength of attention across signal frequencies for the three conditions (no cue, signal cue, masker cue) reported above. The fidelity with which different frequencies regions were evaluated was modest compared to the earlier studies. The current approach, however, used only no-signal trials, which has the advantage that trials are drawn from a single distribution across conditions.

To clarify the intent of the analysis, imagine an observer is able to remember the five different signal frequencies used. What errors might such an observer make? Considering just the no-signal trials, when a masker component falls at or near one of the signal frequencies, false-alarm rates should be higher than when none of the masker components falls near the signal frequencies. By comparing observer's responses with the frequency composition of the stimuli presented on each trial, consistent patterns of responses can be revealed.

1. Computational methods

A common method of evaluating trial-by-trial data is to fit a linear model. For example, it might be assumed that a decision variable is formed based on weighted and summed levels at the outputs of independent auditory filters (Ahumada and Lovell, 1971; Berg, 1989; Richards and Zhu, 1994; Lutfi, 1995). These models are best applied when the independent variables (level at the output of auditory filters in the above example) are normally distributed, although the linear models appear to be robust to violations of such assumptions. Here, a d'-like analysis was used to evaluate observers' strategies.

Only no-signal trials were considered, and so one need only consider a representation for the masker. We used a power spectrum-like representation, that is, temporal interactions were ignored. The masker's power spectrum was zero-valued, except for the frequencies occupied by the six 50-dB SPL masker components. An equivalent description is a vector of ones and zeros, where each entry indicates whether or

not there was energy at a particular frequency. For any one element of the vector, across trials the distribution is approximately Bernoulli. The length of the vector (number of vector elements) depends on frequency gradation—in the current analysis 19 bins tile the frequency range from 200 to 5000 Hz. The bin size corresponded approximately to the size of the protected region (±8%, except below 500 Hz[3]) and five of the bins were centered at the five signal frequencies. On some trials, two or more maskers fell into one bin, leading to a level greater than 50 dB SPL. Even so, the bin was assigned a value of 1. The reasoning was that the change from 50 (e.g., to 53 dB SPL) was small compared to the change from no energy to a level of 50 (or 53) dB SPL.

The data analyzed were lists of 19-element vectors filled with zeros and ones and a list of an observer's responses, also coded using zeros (''no signal'') and ones (''signal''). The length of the list was the number of trials. For each frequency bin (vector element), across all trials a two-by-two table was generated whose entries were in essence hits, false alarms, etc. Note that these values were derived using only the no-signal stimuli, and so have a very different meaning than is typical in signal detection theory applications. As an example, consider the construction of the two-by-two table for the lowest-frequency bin. When there was energy present in the bin (due to a masker component randomly falling into the bin) and the observer responded ''signal,'' the trial was counted as a hit. When the observer responded ''signal'' but there was no energy in the bin, the trial was counted as a false alarm. The other cells of the two-by-two (misses and correct rejections) could be similarly constructed. For each frequency bin, hit and false-alarm rates were thus generated, and the terms hit_E and FA_E denote these values. Finally, for each frequency bin a value labeled d_E was determined, which is the observer's sensitivity to the presence of energy at a particular frequency region as signifying the presence of a signal. The subscript ''E'' is a reminder that the analysis was relative to *energy* in a bin, and d is used instead of d' to emphasize that the analysis is relative to energy in the bin, not the presence of an external *signal*. d_E is defined as d' would be for a YN task: $d_E = z(\text{hit}_E) - z(\text{FA}_E)$, where z is the standard normal score (Macmillan and Creelman 1991). Applying this procedure across all frequency bins yielded plots of sensitivity to energy as a function of frequency, which are referred to as sensitivity functions. Both simulations and reanalysis of the data indicated that fits to a linear model provided nearly identical patterns of ''relative weights'' as a function of frequency.[5]

For each observer, all no-signal trials from the no-cue condition and from the masker-cue conditions contributed to the sensitivity functions. For the signal-cue condition, data from ISIs of 5 ms were not included.

2. Sensitivity functions and discussion

Figure 6 shows sensitivity functions averaged across observers. Averaged sensitivity functions provide a rough description of group strategies rather than strategies for individual observers. The top, middle, and bottom panels are for the signal-cue, masker-cue, and no-cue conditions, respectively. Error bars are standard errors of the mean across observers and the vertical dashed lines indicate the five signal frequencies.

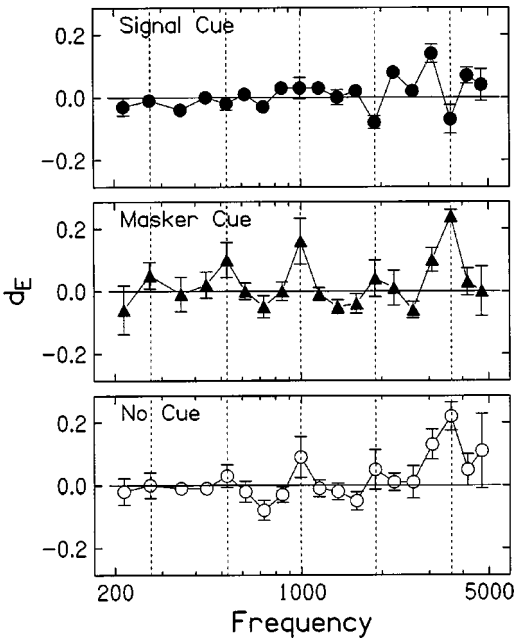


FIG. 6. Averaged sensitivity functions are plotted for the signal-cue, masker-cue, and no-cue conditions (top to bottom). Error bars indicate the standard errors of the mean across observers (six observers). Vertical dashed lines mark the signal frequencies.

The averaged sensitivity functions for the no-cue and masker-cue conditions were highly correlated [$r = 0.87$, $F(1,18) = 51.76$, $p < 0.001$]. In contrast, neither the no-cue nor the masker-cue sensitivity functions were correlated with the averaged sensitivity function derived for the signal-cue condition ($r < 0.01$ and $r = 0.01$, respectively). This indicated that at least some of the observers adopted different strategies in the signal-cue condition compared to the other two conditions.

For the high frequencies in the no-cue condition (bottom panel) d_E was elevated in the region of the signal frequencies and depressed between the signal frequencies. This elevation was clearest for the 995- and 3630-Hz signals, and indicated that energy in the region of these signal frequencies elicited ''signal'' rather than ''no-signal'' responses (and/or a lack of energy elicited ''no signal'' responses). It is apparent that observers both remembered which frequencies were potential signal frequencies and attended to those frequencies (see Buus *et al.*, 1986, for similar conclusions based on results using multitonal signals). When there was no cue, observers appear to have paid somewhat more attention to a portion of the spectrum, i.e., the higher signal frequencies. Thus, even though thresholds measured in the no-cue condition exceeded the ''loudness limit'' (Fig. 5), the current analysis suggests that not all observers rely only on changes in loudness to detect the added tone when the added tone was randomly chosen from among five.

The peaks in the sensitivity functions at potential signal frequencies were even more apparent in the masker-cue condition (middle panel) compared to the no-cue condition (bottom panel). It seems reasonable to assume that observers

adopted similar strategies in the two conditions, or at least the strategies shared features. If so, then how should one interpret the improvement in threshold when masker cues are added? Perhaps the masker preview provided a referent frequency map for the observers, thereby allowing a more accurate ''attention window'' (Dai *et al.*, 1991) in the region of the potential signal frequencies. Alternately, because the frequency was the same in the cue and the trial, comparing the cue and trial stimuli might result in less variance in a decision variable than when there was no cue (i.e., a low-variance standard is available). The reduced variance, in turn, might lead to reduced thresholds (e.g., Lutfi, 1993; Oh and Lutfi, 1999; Tang and Richards, 2003).[6]

Lower thresholds in the masker-cue condition than in the no-cue condition are consistent with the observation that in a 2IFC task, thresholds were lower when the maskers were the same vs different across observation intervals (e.g., FC-F vs FC-R in Sec. II; see also Neff and Callaghan, 1988; Neff and Dethlefs, 1995; Wright and Saberi, 1999; Tang and Richards, 2003). The sensitivity functions add to this observation in suggesting that observers' basic strategies may not depend heavily upon whether or not information about the masker is provided.

The shape of the sensitivity functions obtained in the no-cue and masker-cue conditions also allows potential detection models to be rejected. In neither condition did observers appear to compare the energy in the cue vs trial interval. Such a strategy would yield a flat sensitivity function. Nor does it seem likely that the masker cue acted to ''potentiate'' the frequencies in the region of the masker components (exogenous cuing; e.g., Hafter *et al.*, 1993; Green and McKeown, 2001). A simple priming effect would lead to lower thresholds in the no-cue than the masker-cue conditions, and sensitivity functions associated with the masker cue would be relatively flat and near zero, not an exaggerated version of the no-cue function.

Finally, consider the sensitivity function for the signal-cue condition (top panel). It is important to keep in mind that: (a) the functions are derived from no-signal trials; and (b) the masker did not have energy in the region of the signal frequency, i.e., the protected region. As a result, d_E values cannot be estimated at the signal frequency. The current method can, however, detect the *absence* of sensitivity at the other four signal frequencies. It is apparent that compared to the other two conditions, there were no peaks at the higher signal frequencies for the signal-cue condition. This implies that observers used the signal cue to shift their attention to the appropriate (cued) signal frequency and away from the other potential (not-cued) signal frequencies.

V. SUMMARY AND CONCLUSIONS

The results of the first series of experiments showed that a signal cue reduced thresholds in three conditions: (1) a YN procedure in which masker frequencies were randomized every trial; (2) a 2IFC procedure in which the maskers were randomly drawn prior to each interval (FC-R); and (3) a 2IFC procedure in which the maskers were randomly drawn prior to a trial, but fixed across intervals (FC-F). This basic result held for maskers comprised of two, six (FC-R condition only), and ten components. Individual differences in the magnitude of the cuing effect were largest when two-component maskers were tested. The magnitude of the cuing effect was not simply related to thresholds measured without a cue. Thus, the poorest performers do not necessarily benefit most from the cue.

The second series of experiments showed that a signal cue provided substantial release from masking when the signal frequency was chosen at random. In these conditions observers' thresholds were quite high, possibly reflecting a dependence on changes in loudness to detect the added tone. If both signal frequency and overall stimulus level were varied, most listeners could not perform above chance unless provided with a pretrial cue to the signal frequency.

The third series of experiments showed that varying ISI had relatively little effect on the threshold reductions produced by both masker and signal cues. The signals were random, drawn from five potential frequencies. The only condition that did not produce an improvement in performance with the cue was the shortest ISI (5 ms) for the signal cue, perhaps because of some form of forward masking of the signal by the preceding cue. Possibly due to ''edge'' effects for the range of potential signal and masker frequencies, performance was best for the highest signal frequency for both the masker- and signal-cue conditions, but not for the no-cue condition.

An analysis of the data from the third experiment derived sensitivity functions to assess observers' decision strategies. The patterns of the averaged sensitivity functions were similar for the no-cue and masker-cue conditions, but different for the signal-cue condition. For the no-cue and masker-cue conditions, the functions showed peaks at several potential signal frequencies, indicating listeners both remembered and attended to these frequencies. The largest peaks were observed for the highest frequency signals. Such peaks were not observed for the sensitivity functions for signal-cue conditions, suggesting that listeners relied on the cue to indicate the signal frequency for the upcoming trial.

Overall, the data indicate that pretrial cues to either the signal or the masker can improve performance with random-frequency maskers for the majority of listeners. This holds whether the signal frequency is constant or drawn at random on different trails, although the release in masking is larger for the latter conditions. The pattern of thresholds suggests that listeners do not rely heavily on loudness as a detection strategy unless the signal frequency is drawn at random. Even then, trial-by-trial analyses suggest that observers do not rely completely on differences in loudness. If best performance is desired for masking studies with frequency uncertainty, the present results indicate that a 2IFC procedure should be used with a signal cue.

ACKNOWLEDGMENTS

We are indebted to Zhongzhou Tang, who motivated the d' analysis of the data in experiment III. We thank Brian Callaghan, Rong Huang, and Fanfan Xiong for assistance with data collection and analysis. Dr. Soren Buus and an anonymous reviewer provided very useful suggestions on an

earlier version of this manuscript. This work was supported by grants R01 DC00925 (BTNRH) and RO1 DC02012 (UPenn) from the National Institutes of Health.

[1]The choice of using *t*-tests rather than analysis of variance was due to the use of different observers for $N=6$ compared to $N=2$ and 10.

[2]Depending on presumed criterion placement, there are a range of d' values that might lead to 71% correct decisions in a YN procedure. Using standard signal-detection theory analyses, and assuming equal variance signal and no-signal distributions, the range of d' values can be reasonably bracketed by two criterion placements. If the criterion is placed so as to maximize percent correct, $d'=1.1$ will lead to 71% correct decisions. If, on the other hand, the criterion is placed so that the false-alarm rate is 0.5, $d'=1.4$ will provide 71% correct detections. Thus, one would expect the d' value associated with the track-based threshold to be slightly higher than 1.

[3]If the signal frequency was equal to or below 500 Hz, the protected region was fixed at ± 40 Hz.

[4]Thresholds were somewhat higher for the highest signal frequency. As a result, it seemed prudent to estimate false-alarm rates separately for the different signal frequencies and levels. Spot checks indicate this step was probably unnecessary; threshold estimates are essentially the same whether separate or aggregate false-alarm rates are tested.

[5]In particular, for the current data set weights associated with correlation coefficients (Φ coefficient, r_Φ) and a linear model provided results nearly identical to the sensitivity functions described in the text.

[6]Whether or not the differencing operation leads to reduced variance across trials depends on how the energy is combined across frequency. For example, if level estimates are accumulated across all frequencies, including those frequencies with no energy, differencing the decision variables for the cue and trial stimuli might lead to higher thresholds than using the decision variable for the trial stimulus (depending on the magnitude of memory variability, etc.).

Ahumada, Jr., A., and Lovell, J. (**1971**). ''Stimulus features in signal detection,'' J. Acoust. Soc. Am. **49**, 1751–1756.

Berg, B. G. (**1989**). ''Analysis of weights in multiple observation tasks,'' J. Acoust. Soc. Am. **86**, 1743–1746.

Buus, S., Schorer, E., Florentine, M., and Zwicker, E. (**1986**). ''Decision rules in detection of simple and complex tones,'' J. Acoust. Soc. Am. **80**, 1646–1657.

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.

Dai, H., Scharf, B., and Buus, S. (**1991**). ''Effective attenuation of signals in noise under focused attention,'' J. Acoust. Soc. Am. **89**, 2837–2842.

Green, D. M. (**1961**). ''Detection of auditory sinusoids of uncertain frequency,'' J. Acoust. Soc. Am. **33**, 897–903.

Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford University Press, Oxford).

Green, T. J., and McKeown, J. D. (**2001**). ''Capture of attention in selective frequency listening,'' J. Exp. Psychol. Human Percept. Perform. **27**, 1197–1210.

Hafter, E. R., Schlauch, R. S., and Tang, J. (**1993**). ''Attending to auditory filters that were not stimulated directly,'' J. Acoust. Soc. Am. **94**, 743–747.

Kidd, G., Jr., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). ''Reducing informational masking by sound segregation,'' J. Acoust. Soc. Am. **95**, 3475–3480.

Kidd, G., Jr., Mason, C., and Arbogast, T. (**2002a**). ''Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns,'' J. Acoust. Soc. Am. **111**, 1367–1376.

Kidd, G., Jr., Mason, C. R., and Richards, V. M. (**2002b**). ''Measuring informational masking using the multiple-bursts procedure: Effect of number of bursts and interburst interval,'' Abstracts of the 25th Midwinter Meeting, Assoc. Research in Otolar, **25**, p. 179(A).

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

Lutfi, R. A. (**1993**). ''A model of auditory pattern analysis based on component-relative-entropy,'' J. Acoust. Soc. Am. **94**, 748–758.

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.

Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge).

Moore, B. C. J. (**1981**). ''Interactions of masker bandwidth with signal duration and delay in forward masking,'' J. Acoust. Soc. Am. **70**, 62–68.

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.

Neff, D. L. (**1991**). ''Forward masking by maskers of uncertainty frequency content,'' J. Acoust. Soc. Am. **89**, 1314–1323.

Neff, D. L., and Callaghan, B. P. (**1988**). ''Effective properties of multicomponent simultaneous maskers under conditions of uncertainty,'' J. Acoust. Soc. Am. **83**, 1833–1838.

Neff, D. L., and Green, D. M. (**1987**). ''Masking produced by spectral uncertainty with multicomponent maskers,'' Percept. Psychophys. **41**, 408–415.

Neff, D. L., and Dethlefs, D. L. (**1995**). ''Individual differences in simultaneous masking with random-frequency, multicomponent maskers,'' J. Acoust. Soc. Am. **98**, 125–134.

Oh, E. L., and Lutfi, R. A. (**1999**). ''Nonmonotonicity of informational masking,'' J. Acoust. Soc. Am. **106**, 3521–3528.

Oh, E. L., Wightman, F., and Lutfi, R. A. (**2001**). ''Children's detection of pure-tone signals with random multitone maskers,'' J. Acoust. Soc. Am. **109**, 2888–2895.

Penner, M. J. (**1972**). ''The effect of payoffs and cue tones on detection of sinusoids of uncertain frequency,'' Percept. Psychophys. **11**, 198–201.

Press, W. H., Flannery, B. R., Teukolsky, S. A., and Vettering, W. T. (**1986**). *Numerical Recipes: The Art of Scientific Computing* (Cambridge Press, Cambridge).

Raney, J. J., Richards, V. M., Onsan, Z. A., and Green, D. M. (**1989**). ''Signal uncertainty and psychometric functions in profile analysis,'' J. Acoust. Soc. Am. **86**, 954–960.

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.

Schlauch, R. S., and Hafter, E. R. (**1991**). ''Listening bandwidths for frequency uncertainty in pure-tone signal detection,'' J. Acoust. Soc. Am. **90**, 1332–1339.

Spiegel, M. F., Picardi, M. C., and Green, D. M. (**1981**). ''Signal and masker uncertainty in intensity discrimination,'' J. Acoust. Soc. Am. **70**, 1015–1019.

Tang, Z., and Richards, V. M. (**2003**). ''Examination of a linear model in an informational masking study,'' J. Acoust. Soc. Am. **114**, 361–367.

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.

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.

Zwicker, E. (**1984**). ''Dependence of post-masking on masker duration and its relation to temporal effects in loudness,'' J. Acoust. Soc. Am. **75**, 219–223.

Within-ear and across-ear interference in a dichotic cocktail party listening task: Effects of masker uncertainty

Douglas S. Brungart[a] and Brian D. Simpson
Air Force Research Laboratory, 2610 Seventh Street, Wright-Patterson AFB, Ohio 45433-7901

(Received 3 February 2003; revised 22 September 2003; accepted 23 September 2003)

Increases in masker variability have been shown to increase the effects of informational masking in non-speech listening tasks, but relatively little is known about the influence that masker uncertainty has on the informational components of speech-on-speech masking. In this experiment, listeners were asked to extract information from a target phrase that was presented in their right ear while ignoring masking phrases that were presented in the same ear as the target phrase and in the ear opposite the target phrase. The level of masker uncertainty was varied by holding constant or ''freezing'' the talkers speaking the masking phrases, the semantic content used in the masking phrases, or both the talkers and the semantic content in the masking phrases within each block of 120 trials. The results showed that freezing the semantic content of the masking phrase in the target ear was the only reduction in masker uncertainty that ever resulted in a significant improvement in performance. Providing feedback after each trial improved performance overall, but did not prevent the listeners from making incorrect responses that matched the content of the frozen target-ear masking phrase. However, removing the target-ear contents corresponding to the masking phrase from the response set resulted in a dramatic improvement in performance. This suggests that the listeners were generally able to understand both of the phrases presented to the target ear, and that their incorrect responses in the task were almost entirely a result of their inability to determine which words were spoken by the target talker. [DOI: 10.1121/1.1628683]

PACS numbers: 43.66.Pn, 43.66.Rq, 43.71.Gv [LRB] Pages: 301–310

I. INTRODUCTION

Because multitalker listening plays such an important role in interpersonal communication, a great deal of research has been devoted to studying the extraction of information from target speech signals that are masked by one or more interfering talkers [see Bronkhorst (2000) and Ericson and McKinley (1997) for recent reviews on the subject]. However, many questions related to multitalker listening remain unanswered. One aspect of multitalker speech perception that has thus far received little attention is the influence that *a priori* knowledge about the characteristics or content of the interfering speech waveform has on a listener's ability to extract information from the target speech signal. Although multitalker speech perception studies have been conducted with interfering speech samples that were randomly selected in each trial (Egan *et al.*, 1954; Abouchacra *et al.*, 1997; Brungart, 2001; Drullman and Bronkhorst, 2000), interfering speech samples that were repeated in a predictable order (Freyman *et al.*, 1999), and interfering speech samples with content that was known prior to each trial of the experiment (Hawley *et al.*, 1999; Darwin and Hukin, 2000), none of these studies have explicitly examined how the intelligibility of the target speech signal was influenced by the predictability of the interfering speech. However, the results of experiments examining the role of ''informational masking'' in the processing of complex auditory stimuli suggest that reductions in masker uncertainty are likely to result in some performance improvements in multitalker listening tasks.

Previous research has shown that two distinct kinds of masking may impact performance in these multitalker listening tasks. The first is ''energetic'' masking, which occurs when the target and masking waveforms overlap in time and frequency in such a way that portions of the target signal are rendered undetectable. The second type of masking is ''informational'' masking, which occurs when target and masking waveforms that do not overlap in time and frequency are perceptually ''similar'' in such a way that the listener is unable to distinguish between the acoustic elements associated with the target signal and the acoustic elements associated with the masking signal (Brungart, 2001; Brungart *et al.*, 2001; Freyman *et al.*, 1999; Kidd *et al.*, 1995). In multitalker listening, informational masking could occur either at the phonetic level, where the listener might have difficulty segregating the basic acoustic elements of the two speech signals into coherent syllables and words, or it could occur at the semantic level, where the listener might hear two distinct coherent words and be unable to determine which word was spoken by which talker.

While a reduction in masker uncertainty is unlikely to provide much release from the energetic components of speech-on-speech masking, it has the potential to produce a substantial reduction in the informational components of speech-on-speech masking. At the phonetic level of informational masking, the potential advantages of reduced masker uncertainty might be analagous to the ones that have been demonstrated in nonspeech informational masking experiments where listeners were asked to detect specific target sounds in the context of nonenergetic maskers that were randomly selected from progressively smaller sets of potential

[a]Electronic mail: douglas.brungart@wpafb.af.mil

waveforms. For example, Watson and his colleagues measured the ability of listeners to detect a small change in frequency in one of ten short consecutive tones (Watson *et al.*, 1976), and found that the smallest detectable relative change in frequency ($\Delta f/f$) decreased from 14% when the 10-tone pattern was randomly selected from a catalog of 50 possible patterns in each trial to 1.4% when the same pattern was used in every trial. More recently, Wright and Saberi (1999) reported that the detection threshold for a fixed-frequency tone masked by a random-frequency masker decreased by an average of 9 dB as the number of possible masking patterns was reduced from ten to two. Even larger reductions in detection threshold can occur when listeners are given explicit *a priori* information about the characteristics of the random-frequency masking signal prior to the presentation of the stimulus. Richards and Tang (2002), for example, reported a 15-dB reduction in detection threshold in a multitone informational masking task when subjects were given a pre-cue of the masker prior to each stimulus presentation. These results show that reductions in masker uncertainty can lead to improvements in a listener's ability to extract information from complex acoustic stimuli that may make it easier to attend to a target speech signal masked by one or more interfering talkers.

Evidence that masker uncertainty might influence the amount of informational masking that occurs at the semantic level in multitalker listening tasks is provided by a previous experiment (Brungart, 2001) that showed that listeners in a two-talker color and number identification task were much more likely to make incorrect responses that matched the content of the masking speech than they were to make incorrect color and number responses that were not spoken by either of the talkers in the stimulus. It was hypothesized that this occurred because the listeners understood both color-number pairs in the stimulus but were unable to correctly determine which one was spoken by the target talker. If the listeners were indeed hearing both color-number pairs in this experiment, then one might expect them to be able to use elimination to correctly identify the target color-number pair if they had *a priori* information about the color and number coordinates contained in the masking phrase.

These kinds of elimination-based strategies might be even more powerful in dichotic listening situations where informational masking is increased by the presence of an interfering talker in the unattended ear. In a recent experiment, we examined situations where a target talker and a masking talker were mixed together and presented to one ear and a second independent masking talker was presented to the other ear. The results showed that listeners were substantially more likely to confuse the content of the target phrase with the content of the target-ear masking phrase when a second masking phrase was presented to the unattended ear than when a nonspeech masker or no masker was presented to the unattended ear (Brungart and Simpson, 2002). Because listeners are more likely to make target-masker confusions in dichotic listening configurations with a contralateral masker, one might expect them to obtain more benefit from *a priori* information about the target-ear masking phrase in these configurations than in diotic or monaural configurations where target-masker confusions are less prevalent. Also of interest is how the degree of uncertainty of the masking phrase in the unattended ear affects performance in these kinds of dichotic listening tasks. In this paper, we describe the results of three experiments that have examined the relation between the level of uncertainty associated with a masking speech phrase and the degree to which that masking phrase causes informational masking in a monaural or dichotic multitalker listening task.

II. EXPERIMENT 1: MASKER UNCERTAINTY

A. Methods

1. Listeners

Seven paid volunteer listeners, four male and three female, participated in the experiment. All had normal hearing ($<$15 dB HL from 500 Hz to 8 kHz), and their ages ranged from 21 to 55 years. All of the listeners had participated in previous experiments that utilized the speech materials used in this study.

2. Stimuli

a) Speech materials. The speech stimuli were taken from the publicly available Coordinate Response Measure (CRM) speech corpus for multitalker communications research (Bolia *et al.*, 2000). This corpus consists of phrases of the form ''Ready (call sign) go to (color) (number) now'' spoken with all possible combinations of eight call signs (''arrow,'' ''baron,'' ''charlie,'' ''eagle,'' ''hopper,'' ''laker,'' ''ringo,'' ''tiger''), four colors (''blue,'' ''green,'' ''red,'' ''white''), and eight numbers (1–8). Thus, a typical utterance in the corpus would be ''Ready baron go to blue five now.'' Eight talkers (four male, four female) were used to record each of the 256 possible phrases, so a total of 2048 phrases are available in the corpus. Variations in speaking rate were minimized by instructing the talkers to match the pace of an example CRM phrase that was played prior to each recording. The sentences in the corpus, which are band-limited to 8 kHz, were resampled from the original 40-kHz sampling rate to 25 kHz to reduce computation time in the processing of the stimuli. The phrases were time aligned to ensure that the word ''ready'' started at the same time in all the speech signals in the stimulus, but no additional efforts were made to synchronize the call signs, colors, and numbers in the competing CRM phrases.

b) Target speech signal. In each trial of the experiment, the target speech signal was randomly selected from the 32 phrases spoken by the target talker with the call sign ''Baron.'' In the first half of the experiment, the target talker was always the same male talker (talker 0 from the CRM corpus). In the second half of the experiment, the target talker was always the same female talker (talker 4 from the CRM corpus). In both cases, the listener was familiarized with the voice of the target talker prior to the collection of the first block of data. The target speech signal was always presented monaurally to the listener's right ear.

c) Target-ear masking signal. In both the monaural and dichotic conditions of the experiment, the target signal presented to the listener's right ear was electronically mixed with a masking phrase that was selected from the 441 phrases in the corpus that contained a different call sign, a different color, and a different number than the target phrase and were spoken by a different talker who was the same sex as the target talker (7 call signs×3 colors×7 numbers×3 same-sex talkers). The level of this masking signal was randomly varied over a 6-dB range in each trial of the experiment, and the rms value of the target signal was scaled relative to the randomly selected rms value of the masker to set the signal-to-noise ratio (SNR) in the target ear to one of five different values (−8, −4, 0, +4, or +8 dB).

d) Contralateral-ear masking signal. In the dichotic conditions of the experiment, a second masking phrase was selected from the 144 phrases in the corpus with a different same-sex talker, a different call sign, a different color, and a different number than either the target phrase or the target-ear masking phrase (6 call signs×2 colors×6 numbers×2 same-sex talkers). This masking phrase was presented monaurally to the listener's left ear at the same overall rms level as the masking phrase in the listener's right ear.

3. Masker uncertainty

Four different levels of masker uncertainty were used in the experiment: *random*, where the masker in each trial was selected randomly from all the possible masking phrases in the corpus; *talker frozen*, where a single masking talker was randomly selected prior to each block of 120 trials and the masking phrase in each trial was selected randomly from the CRM phrases spoken by the preselected masking talker; *content frozen*, where a single call sign, color, and number combination was chosen randomly prior to each block of trials and the masking phrase in each trial was selected from the phrases in the corpus containing the preselected combination of call sign, color, and number; and *both frozen*, where a single masker phrase was selected randomly prior to each block of 120 trials and that same phrase was used as the masker in each trial in that block.

These four different levels of masker uncertainty were applied to the target ear, the contralateral ear, or both ears in one of the 14 different stimulus conditions illustrated in Table I. The first two conditions were monaural and dichotic baseline conditions where the masking phrases were selected randomly. Conditions 3–5 were monaural conditions with no masker in the contralateral ear. Conditions 6–8 were dichotic conditions where only the masker in the target ear was frozen. Conditions 9–11 were dichotic conditions where only the masker in the contralateral ear was frozen. Conditions 12–14 were dichotic conditions where the same level of masker uncertainty was applied to both ears. Note that no dichotic conditions were conducted with different levels of masker uncertainty in the two ears.

4. Procedure

The data were collected with the listeners seated in front of the CRT of a Windows-based control computer in a quiet, sound-treated listening room. The stimuli for each trial were generated by an interactive MATLAB script, which selected the stimulus signals from the CRM corpus, electronically mixed the target and masking signals for the left and right ears, and presented the signals over headphones (AKG K-240DF) at a comfortable listening level (roughly 70 dB SPL) through a D/A converter (Tucker-Davis DD1) connected to a stereo headphone amplifier (Tucker-Davis HB6). The listeners were instructed to listen in their right ear for the target phrase, which was always addressed to the call sign "Baron," and to use the mouse to select the color and number contained in the target phrase from an array of colored digits displayed on the screen of the control computer. Each listener first participated in 3 blocks of 120 trials in each of the 14 conditions shown in Table I with the male target talker. They then participated in an additional 3 blocks of 120 trials in each of the 14 conditions with the female target talker. Thus each of the 7 listeners participated in a total of 84 blocks of 120 trials at a rate of 1–3 blocks per day over a period of roughly 10 weeks. This resulted in a total of 70 560 trials, or 1008 trials at each of the five target-ear SNR values in each of the 14 conditions of the experiment.

TABLE I. Masker uncertainty conditions.

			Experiment no.		
Condition	Target ear	Unattended ear	1	2a	2b
1	Random	No masker	√	√	
2	Random	Random	√	√	
3	Talker frozen	No masker	√		
4	Content frozen	No masker	√	√	
5	Both frozen	No masker	√		
6	Talker frozen	Random	√		
7	Content frozen	Random	√	√	√
8	Both frozen	Random	√		
9	Random	Talker frozen	√		
10	Random	Content frozen	√	√	√
11	Random	Both frozen	√		
12	Talker frozen	Talker frozen	√		
13	Content frozen	Content frozen	√	√	√
14	Both frozen	Both frozen	√		

5. Training

Because the experiment was based on the premise that the listeners would be able to acquire knowledge about the statistical properties of the masking phrases in each block of trials and then formulate response strategies based on this knowledge, some of the trials at the beginning of each block had to be counted as training trials and discarded from the analysis of the data. An analysis of performance as a function of trial sequence conducted after the completion of the experiment indicated that the percentage of correct responses increased rapidly for the first 10–15 trials of each block and then plateaued for the remaining 105–110 trials. Thus, the first 20 trials of each block were eliminated from all of the analyses presented in the following sections.

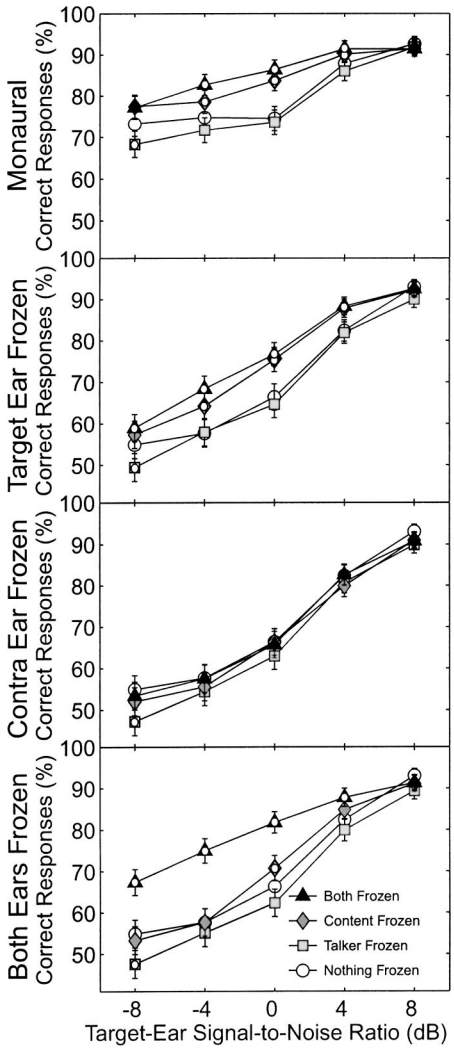


FIG. 1. Percentages of correct color-number responses as a function of target-ear SNR value for each of the 14 different conditions tested in the experiment. The different symbols represent different levels of masker uncertainty, as indicated by the legend. The top panel shows performance in the monaural configuration, and the bottom three panels show performance in the dichotic configurations where the masking phrases were frozen in the target ear, the contralateral ear, or in both ears, respectively. The error bars show 95% confidence intervals calculated from the raw data for each data point, and the white dots on the boldly outlined symbols indicate points that were significantly different from the random baseline condition at the $p<0.05$ level (Fisher LSD post-hoc analysis).

B. Results and discussion

Figure 1 shows overall performance as a function of target-ear SNR in each of the 14 conditions tested in the experiment. Each data point in the figure represents the percentage of trials in which the listeners correctly identified both the color and number in the target phrase. These data have been plotted as a function of target-ear SNR, with a different symbol for each level of masker uncertainty (as indicated by the legend). The four panels represent four different configurations of the frozen and random maskers: the top panel shows performance in the monaural configuration where no contralateral masker was present; the second panel shows performance in the dichotic configurations where only the masking talker in the target ear was frozen; the third panel shows performance in the dichotic configuration where only the masker in the contralateral ear was frozen; and the bottom panel shows performance in the dichotic configuration where the maskers in both ears were frozen. Note that the random dichotic condition has been plotted as a reference in each of the bottom three panels of the figure (open circles). Also note that the error bars show the 95% confidence intervals of each data point, calculated from the raw data from each condition of the experiment.

These data were also analyzed with a two-factor within-subjects repeated-measures ANOVA, with condition number (1–14) and target-ear SNR (−8, −4, 0, +4, or +8 dB) as the two independent variables, and with the arcsine-transformed proportion of correct responses at a given target-ear SNR value within a single block of trials treated as one of six independent replications of each condition for each listener. The results of this ANOVA indicate that there were significant main effects for both target-ear SNR ($F_{(4,24)}=22.31$, $p<0.00001$) and condition number ($F_{(13,78)}=8.95$, $p<0.00001$), and that there was a significant interaction between these two factors ($F_{(52,312)}=4.89$, $p<0.00001$). A *post hoc* test (Fisher LSD) was performed on this interaction to determine which points in Fig. 1 were significantly different (at the $p<0.05$ level) from the corresponding random-masker control conditions with the same target-ear SNR values. Small white dots and bold outlines have been used to mark the symbols in the figure that correspond to conditions that were significantly different than the random-masker control condition.

A number of important observations can be made from the results shown in the figure:

(1) In the baseline conditions with random masking phrases, the results were consistent with those of earlier experiments that have used similar stimuli. In the monaural baseline condition (open circles in the top panel of Fig. 1), performance decreased systematically as the target-ear SNR decreased from +8 to 0 dB, then plateaued at approximately 75% correct responses at target-ear SNR values less than 0 dB. In the dichotic baseline condition (open circles in the bottom three panels of the figure), performance was similar to the monaural configuration when the target-ear SNR was greater than 0 dB, but was significantly worse than the monaural configuration when the SNR was less than or equal to 0 dB (Fisher LSD, $p<0.05$). Thus it appears that the presence of the masking speech signal in the contralateral ear made it significantly more difficult for the listeners to extract information from the less intense talker in the target ear. This same pattern of performance was found in an earlier experiment (Brungart and Simpson, 2002) that examined within-ear and across-ear interference with the same stimuli used in this experiment.

(2) Freezing the content of the target-ear masking phrase had a much greater impact on performance than any of the other masker uncertainty manipulations tested in this experiment. Significant improvements relative to the random-masker baseline conditions occurred in every condition where the content of the target-ear masking phrase was frozen (as indicated by the white dots on the

diamonds and triangles in the first, second, and last panels of Fig. 1), and never occurred in any condition where the content of the target-ear masking phrase was not frozen.

(3) Freezing only the masking talkers never produced any improvement in performance, and it actually caused a significant decrease in performance in every configuration when the target-ear SNR was −8 dB (squares in Fig. 1).[1] Freezing the masking talkers in conjunction with the content of the masking phrases (triangles in Fig. 1) produced a significant improvement in performance relative to the content-frozen condition (diamonds in the figure) only when the masking phrases were frozen in both ears (bottom panel of the figure) and the target ear SNR was less than or equal to 0 dB (Fisher LSD, $p < 0.05$).

(4) Freezing the attributes of the contralateral-ear masking phrase also had relatively little impact on performance. No significant improvements ever occurred from freezing just the attributes of the masking phrase in the contralateral ear (third panel of Fig. 1), and freezing the masking phrase in the contralateral ear in conjunction with the masking phrase in the target ear was little different from freezing just the masking phrase in the target ear.[2] The relative unimportance of masker uncertainty in the contralateral ear found in this experiment appears to be consistent with earlier results that have shown that the semantic content of the masking speech signal in the contralateral ear has little impact on the amount of interference it produces in the dichotic CRM listening task (Brungart and Simpson, 2002). In the earlier experiment, meaningless time-reversed speech signals were found to produce just as much across-ear interference as normal CRM phrases when the target-ear SNR was less than 0 dB. If dramatic manipulations like time-reversal have no impact on the amount of interference caused by a contralateral masker in the dichotic CRM listening task, it is not surprising that relatively minor manipulations like a reduction in semantic variability would also fail to produce significant changes in overall performance.

(5) In the configurations where frozen content produced a significant improvement in performance, the improvement was generally greatest at target-ear SNR values near 0 dB. This may be related to the absence of salient intensity-based speech segregation cues at target-ear SNRs near 0 dB. Previous experiments have shown that listeners are able to use differences in level to help segregate similar-sounding talkers in the CRM listening task, and that confusions between the numbers contained in the target and masking phrases are most likely to occur when the two phrases are presented at approximately the same level (Brungart, 2001; Egan *et al.*, 1954). To the extent that learning the contents of a frozen masking phrase can serve as an alternative method for distinguishing between the target and masking talkers in a CRM stimulus, one might expect the effects of this learning to be greatest at SNR values near 0 dB where the level-based segregation cues are weakest.

(6) The performance improvements obtained in the content-frozen and both-frozen conditions of the monaural configuration (top panel of Fig. 1) were almost identical both in pattern and in magnitude to the ones that occurred in the corresponding dichotic configuration of the experiment where only the masker in the target ear was frozen (second panel of Fig. 1). This suggests that the performance advantages that are obtained by reducing the uncertainty of the masking signal in the target ear are effectively independent of the performance disadvantages that occur when a random masking talker is added to the unattended ear.

III. EXPERIMENTS 2a AND 2b: FEEDBACK AND RESPONSE SET LIMITATION

Although freezing the content of the target-ear masker clearly improved performance in experiment 1, the performance benefits obtained in the frozen-content conditions were much smaller than those that theoretically could have occurred if the listeners were taking full advantage of the reduced masker uncertainty in those conditions. Because the masking phrases were restricted from ever containing the same color and number coordinates as the target phrase, the listeners should have been able to improve their performance with an elimination-based listening strategy in which they learned the color-number coordinates in the frozen masking phrases and intentionally refrained from ever making responses that matched those frozen coordinates. There is, however, little evidence that the listeners in this experiment were implementing this kind of elimination-based strategy. In the conditions where the content of the masker was frozen in the target ear, more than 65% of the incorrect number responses and more than 75% of the incorrect color responses matched the color and number coordinates contained in the target-ear masking phrase. This was only slightly better than the 77% and 82% rates at which the responses matched the color and number coordinates of the target-ear masking phrases in the random baseline conditions, and vastly worse than the 0% rates that would have occurred with the ideal elimination-based strategy. While it is true that the listeners were never explicitly told that the target and masking phrases would never contain the same color and number coordinates, one might have expected some of them to be able to learn to use an elimination-based listening strategy after participating in thousands of trials in the experiment. The failure of the listeners to adopt efficient listening strategies in experiment 1 raises two important questions about the role of masker uncertainty in the CRM listening task: (1) whether the listeners would have been better able to take advantage of the reduced uncertainty in the frozen-content conditions if they were provided with feedback about their performance and (2) how good their performance would have been if they had used the ideal elimination-based strategy. In order to address these questions, two follow-on experiments were conducted: one in which the listeners were provided with feedback about the correct response after each trial, and one in which they were effectively forced to adopt the ideal elimination-based listening strategy by prohibiting them from responding with the color and number coordinates contained in the frozen masking phrases.

A. Methods

The procedure used in experiment 2a was essentially identical to the one used in the first experiment with the exception that the listeners were provided with correct-answer feedback after each trial of the experiment. This feedback consisted of a black circle that surrounded the correct colored digit on the CRT of the control computer after each listener response. This circle was turned off prior to the presentation of the next trial.

The procedure used in experiment 2b was identical to the one used in experiment 2a, except that the color and number combinations contained in the frozen-content masking phrases in each block of trials were eliminated from the array of possible color and number responses displayed on the CRT at the end of each trial. This prevented the listeners from ever responding with a color or number value contained in one of the frozen-content masking phrases in a given block of trials.

Only 6 of the 14 conditions in the first experiment were tested in experiment 2a: the two control conditions with completely random masking phrases, and the four experimental conditions where only the content of the masking phrases was frozen. Three of these six conditions (the dichotic configurations with content-frozen maskers) were also tested in experiment 2b. The checkmarks in Table I indicate which conditions were tested in each of the three experiments.

Six of the seven listeners who participated in the first experiment also participated in the second and third experiments. In experiment 2a, each listener first participated in three blocks of 120 trials in each of the six conditions with the male target talker (talker 0 in the CRM corpus) in random order, and then participated in an additional three blocks of 120 trials in each condition with the female target talker (talker 4 in the CRM corpus) in random order. Then, in experiment 2b, each listener participated in two blocks in each of the three conditions with the male target talker followed by two blocks in each of the three conditions with the female target talker. Thus each of the six listeners participated in a total of 4320 trials in experiment 2a and 1440 trials in experiment 2b. As in experiment 1, the first 20 trials of each 120 trial block were excluded from the analyses of the data.

B. Results and discussion

1. Overall results

Figures 2 and 3 compare the results for the monaural and dichotic configurations, respectively, tested in experiments 2a and 2b to the corresponding conditions of experiment 1. In each case, the data for the six listeners who participated in all three experiments have been plotted as a function of target-ear SNR with different symbols representing the different possible configurations of the frozen masking phrases (as indicated by the legends). The error bars in the figures represent the 95% confidence intervals calculated from all the raw data at each data point. The data were also analyzed with a two-factor, within-subjects, repeated measures ANOVA, with the five target-ear SNR values (−8, −4, 0, 4, or 8 dB) and the 15 different conditions (six from experiment 1, six from experiment 2a, and three from experiment 2b) serving as the two independent variables, and with the arcsine-transformed proportions of correct responses at each target-ear SNR value within a single block of trials counted as one of four independent replications of each condition for each listener. The results of this ANOVA indicate that there were significant main effects for both target-ear SNR ($F_{(4,20)}=22.59$, $p<0.00001$) and condition number ($F_{(14,70)}=11.12$, $p<0.00001$), and that there was a significant interaction between these two factors ($F_{(56,280)}=3.59$, $p<0.00001$). A *post hoc* test (Fisher LSD) was performed on this interaction to determine which points from experiment 2a were significantly different (at the $p<0.05$ level) from the corresponding conditions with the same target-ear SNR val-

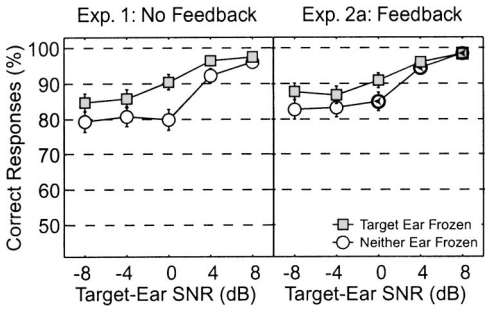


FIG. 2. Percentages of correct color-number responses in the monaural listening conditions for the six listeners who participated in both experiments 1 and 2. The left panel shows performance in experiment 1, where no feedback was provided, and the right panel shows performance in experiment 2a, where the listeners were provided with feedback about the correct color and number combination after each trial. The error bars show 95% confidence intervals calculated from the raw data for each data point, and the arrows on the bold-outlined symbols in the right panel indicate data points in experiment 2a that were significantly different from the corresponding conditions of experiment 1.

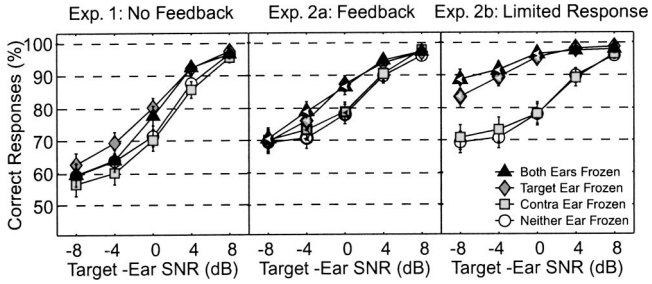


FIG. 3. Percentages of correct color-number responses in the dichotic listening conditions for the six listeners who participated in experiments 1, 2a, and 2b. The left panel shows performance as a function of target-ear SNR from experiment 1, where no feedback was provided. The middle panel shows performance as a function of target-ear SNR for experiment 2a, where the listeners were provided with feedback about the correct color and number combination after each trial. The right panel shows performance as a function of target-ear SNR in experiment 2b, where the listeners were prevented from making responses that matched the colors and numbers contained in the frozen masking phrases. The left-facing arrows on the boldly outlined symbols in the center panel indicate data points in experiment 2a that were significantly different from the corresponding conditions of experiment 1 (Fisher LSD, $p<0.05$). The left-facing arrows on the boldly outlined symbols in the right panel indicate data points in experiment 2b that were significantly different from the corresponding conditions of experiment 2a. Note that the neither ear frozen results from experiment 2a have been replotted in the rightmost panel of the figure to allow an easy comparison with the results of experiment 2b. The error bars show 95% confidence intervals calculated from the raw data for each data point.

ues in experiment 1. The small arrows on the boldly outlined symbols in the right panel of Fig. 2 and in the middle panel of Fig. 3 indicate data points from experiment 2a that were significantly different from the corresponding conditions at the same target-ear SNR in experiment 1. Similarly, the small arrows on the boldly outlined symbols in the right panel of Fig. 3 indicate data points that were significantly different from the corresponding conditions at the same target-ear SNR in experiment 2a. Note that the random-masker baseline from experiment 2a has to be replotted in the right panel of Fig. 3 to allow direct comparison with the data from experiment 2b.

From these results, the following main points can be observed:

(1) The addition of feedback had very little effect on performance in the monaural configurations of the experiment. The only significant effect of feedback was to increase overall performance in the random-masker baseline condition when the target-ear SNR was at least 0 dB (circles in the right panel of Fig. 2).
(2) The addition of feedback produced significant performance improvements in all the dichotic conditions of the experiment when the target-ear SNR was 0 dB or less (center panel of Fig. 3).
(3) Freezing the contents of the target-ear masking phrase produced roughly the same performance improvement relative to the random-masker baseline condition in experiment 2a (comparing the diamonds and triangles to the open circles in the center panel of Fig. 3) as it did in experiment 1 (comparing the same points in the left panel of Fig. 3).
(4) Freezing the contents of the target-ear masking phrase produced much larger benefits in experiment 2b, where the frozen color and number coordinates were eliminated from the response set, than it did in either of the first two experiments (comparing the diamonds and triangles in the three panels of Fig. 3). Indeed, performance in experiment 2b approached 100% correct responses when the target-ear SNR was 0 dB.
(5) Freezing the contents of the contralateral-ear masking phrase, either alone or in conjunction with the contents of the target-ear masking phrase, never had any significant effects on performance in the CRM task.

2. Response distribution

To this point we have discussed the results primarily in terms of the overall percentage of the responses that matched both the color and number in the target phrase. However, additional insights can also be obtained by examining what kind of errors the listeners made when their responses did not match the target phrase. Figure 4 shows how the listeners' responses were distributed in the frozen-content conditions of the three experiments when the target-ear SNR was 0 dB.[3] The responses have been divided into four broad categories: correct responses, where both the color and the number in the response matched the target phrase (shown by the shaded region at the bottom of each bar); within-ear confusions, where the color, the number, or both the color and the number in the response matched the target-ear masker (shown by the white regions of the bars); across-ear confusions, where the color or the number in the response matched the contralateral-ear masker (shown by the black regions of the bars); and responses that did not fall into any of the other three categories (shown by the speckled region at the top of each bar). Responses that included both a within-ear and an across-ear confusion (i.e., included the color from the target-ear masker and the number from the contralateral-ear masker) were extremely rare (never more than 0.4% of the total responses in any condition) and were also grouped into this ''other responses'' category. Note the following important points about these data:

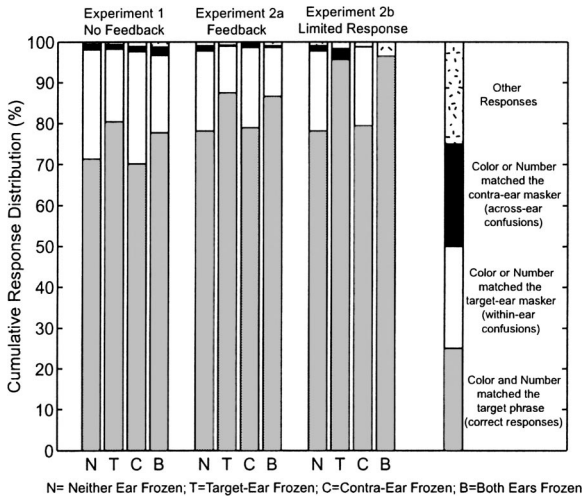


FIG. 4. Cumulative response distributions in the dichotic conditions of the experiments where the target ear SNR was 0 dB and only the content of the masking phrases was frozen. The data from the three experiments have been grouped into three sets of four bars with each set representing a different experiment. Within each set, the four bars represent the random baseline condition where nothing was frozen (*N*), the condition where only the target-ear masking phrase was frozen (*T*), the condition where only the contralateral-ear masking phrase was frozen (*C*), and the condition where both masking phrases were frozen (*B*). Within each bar, the responses have been divided into four broad categories as indicated by the legend. Note that the random baseline *N* condition of experiment 2a has also been plotted as the baseline *N* condition of experiment 2b.

(1) As in earlier experiments that have used the CRM stimuli (Brungart, 2001; Brungart and Simpson, 2002), within-ear confusions accounted for almost all of the incorrect responses in the conditions where listeners were allowed to respond with the color and number coordinates contained in the target-ear masker.
(2) Across-ear confusions accounted for only a very small proportion ($<2\%$) of the incorrect responses in the random-masker baseline conditions of the experiment. This is consistent with the small change in performance that occurred when the content of the contralateral masking phrase was frozen: the listeners almost never made across-ear errors even with random contralateral maskers, so there was little room for improvement from learning to avoid making responses that matched the content of a frozen contralateral-ear masking phrase.
(3) The addition of feedback in experiment 2a resulted in a roughly one-third reduction in the number of within-ear

confusions in every listening configuration of the experiment, including the baseline configuration where nothing was frozen. The feedback also reduced the number of across-ear confusions in the conditions where the masker was frozen in the contralateral ear.

(4) When the frozen color and number coordinates in the target-ear masking phrase were removed from the response set, the eliminated within-ear confusions were almost entirely offset by a corresponding increase in the proportion of correct responses. In the baseline condition of experiment 2a where feedback was provided but none of the attributes of the maskers was frozen (replotted as the *N* condition in the rightmost group of bars in Fig. 4), within-ear confusions accounted for 20% of the total responses. Removing the within-ear masking coordinates from the response set in the target-frozen and both-frozen conditions by definition reduced the number of within-ear confusions to 0%, and in each case this reduction was almost entirely offset by an 18 percentage point increase in the proportion of correct responses. These results suggest that 80%–90% of the within-ear confusions in the random baseline condition of experiment 2a occurred in trials where the listeners actually understood both sets of color-number coordinates and were uncertain about which set was contained in the target phrase, and that only the 10%–20% of the within-ear confusions occurred in trials where the listeners did not correctly perceive the color or number coordinates contained in the target phrase.

IV. GENERAL DISCUSSION AND CONCLUSIONS

The important results of this experiment can be summarized by the following three major points.

Informational masking in the CRM task is dominated by uncertainty about the content of the target-ear masking phrase: This assertion is supported by three major pieces of evidence: (1) Virtually all of the incorrect responses in the experiment matched the content of the target-ear masker. (2) The only reductions in masker uncertainty that ever led to any significant improvements in performance were those where the content was frozen in the target-ear masking phrase. (3) When uncertainty about the target-ear masking phrase was eliminated by removing its contents from the response set, the percentage of incorrect responses at the 0 dB target-ear SNR value decreased from more than 20% incorrect responses to just under 4% incorrect responses. These results provide strong support for the hypothesis that listeners who hear two simultaneous CRM phrases in the same ear are generally able to understand both sets of color and number coordinates and that their incorrect responses are almost exclusively caused by their inability to correctly determine which coordinates were spoken by the target talker and which coordinates were spoken by the masking talker.

None of the other kinds of masker uncertainty examined in this experiment had any meaningful impact on performance in the CRM task: Freezing the talker or content of the masking phrase in the contralateral ear, either alone or in conjunction with the same attributes of the target ear masking phrase, never had any substantial effects on overall performance. Similarly, freezing the masking talker in the target ear, either alone or in conjunction with the contents of the target-ear masking phrase, had very little effect on overall performance. Thus, it appears that the only type of masker variability that has any significant impact on the CRM listening task is uncertainty in the semantic content of the target-ear masking phrase. It is, however, important to note that this result could, in part, be related to the very small and highly redundant color and number response sets used in the CRM speech materials. Intelligibility tasks based on larger vocabularies of phonetically balanced words might be more sensitive to changes in the masker uncertainty other than those directly related to semantic content. Using a single speech waveform repeatedly as the masker might, for example, have a greater effect on a listener's ability to comprehend large sets of nonsense syllables that require accurate perception of every phonetic element than it would on their ability to distinguish between small lists of color words that can be identified from any one of a number of redundant phonetic elements contained within each keyword. It is also important to note that the insensitivity of the CRM task to changes in the uncertainty of masking phrase properties other than semantic content should not be interpreted to mean that the call sign, color, and number have the greatest influence on the masking effectiveness of a particular CRM phrase: in fact, talker has a greater influence than content on the overall masking effectiveness of an individual phrase from the CRM corpus. This is discussed in more detail in the Appendix.

Listeners are relatively poor at developing response strategies that make optimal use of reduced masker uncertainty: Despite having thousands of trials of listening experience, there is no evidence that the listeners in the first two experiments were ever able to effectively adopt the ideal elimination-based listening strategy of learning the color-number coordinates in the frozen target-ear masking phrase and avoiding making responses containing those frozen coordinates. Even in conditions where exactly the same masking waveform was presented to the target ear in every trial and feedback was provided after each incorrect response, performance in experiment 2a never approached the level that was achieved in experiment 2b when the listeners were prohibited from responding with the color and number coordinates contained in the target-ear masking phrase. Indeed, the performance advantages that occurred when feedback was provided applied to all the listening conditions of the experiment equally, and not primarily to the frozen-content conditions as one might expect if the feedback was being used to learn the optimum elimination-based listening strategy. This result is somewhat puzzling. In a typical 120-trial block of experiment 2a with frozen masking content in the target ear, the listeners responded incorrectly 17 times, and 16 of those incorrect responses contained either a color or a number that matched the content of the frozen target-ear masking phrase. Yet, despite the availability of feedback after each trial, the listeners never learned to avoid responding with the coordinates contained in the frozen target-ear masking phrase. A number of factors could have contributed to

this failure to adopt an elimination-based listening strategy. The listeners were never explicitly informed that the target and masking phrases would never have the same color and number coordinates, and this may have led them to respond with the color-number combination contained in the masking phrase even if they recognized that the masking phrase was the same on every trial. Even those listeners who recognized that the masker was frozen and that the target and masking coordinates were never the same may have failed to realize that they could improve their guessing accuracy by never responding with the coordinates contained in the masking phrase. The random ordering of trials with frozen and random maskers may also have made it more difficult to learn an elimination-based listening strategy. A final possibility that is consistent with our own subjective experience with the task is that the listeners may have been so confident that they actually did hear the masking-phrase color or number in the phrase addressed to the target call sign ''Baron'' that they were willing to ignore all of the previous negative feedback and make the incorrect response anyway. Whatever the reason for the failure of the subjects to learn an elimination-based listening strategy in experiment 2a, the results of that experiment suggest that listeners who are engaged in a multitalker listening task may not be particularly good at identifying and exploiting repetitive patterns in the masking phrases used in that task.

In summary, the results of these experiments indicate that there is a direct relation between the variability of the semantic content of a masking phrase and the amount of informational masking it produces in a multitalker listening task. The results of experiment 2b further suggest that specifically instructing listeners not to make responses matching the content of a frozen masking phrase can effectively eliminate all the informational masking that occurs in a restricted response-set task such as the CRM. However, the results of the first two experiments suggest that listeners are unlikely to learn how to take full advantage of reduced semantic masker variability in complex multitalker listening tasks when they are not provided with explicit instructions beforehand. Thus, while there is substantial evidence that masker variability should be taken into account in the design and analysis of multitalker listening experiments, the extent to which masker uncertainty might have a significant impact on speech intelligibility in real-world listening environments is not yet clear.

ACKNOWLEDGMENT

Portions of this work were sponsored by AFOSR LRIR 01-HE-01-COR.

APPENDIX: INTERBLOCK VARIABILITY

An important difference between the random-masker conditions of the experiment and the different frozen-masker conditions of the experiment is that the 120-trial blocks in the frozen conditions contained only a small proportion of the total population of possible masking phrases available in the corpus. In the random conditions, the masking signals in each trial of a block were selected randomly from the 441 same-sex phrases with a different call sign, color, and number than the target phrase (3 talkers×7 call signs×3 colors×7 numbers). In the frozen-talker conditions, the masking signals in each trial of a block were randomly selected from only 147 different phrases (7 call signs×3 colors×7 numbers). In the frozen content conditions, the masking signals in each trial of a block were selected from only three different phrases (three talkers). And in the both-frozen conditions, exactly the same masking phrase was used in every trial of the block.

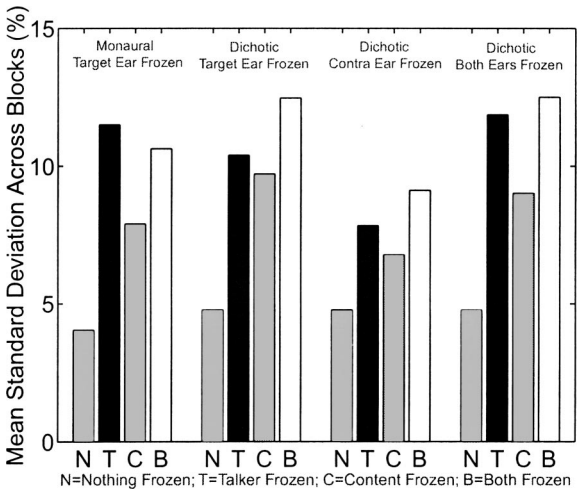


FIG. 5. Mean standard deviation values across the different blocks collected in each listening condition of the experiment. Each listener participated in three different blocks of 120 trials with each of the target talkers in each of the 14 different conditions of the experiment. In this figure, we have calculated the standard deviations of the overall percentages of correct responses across the three blocks in each of the 14 combinations of seven listeners and two target talkers collected in each condition, and averaged these 28 standard deviation values to generate the mean values shown for each condition. Thus the bars show a rough estimate of the variability in performance across different blocks in each of the 14 different conditions of the experiment.

To the extent that there are variations in masking effectiveness across the different phrases in the corpus, one would expect statistical sampling to cause variability in performance across different blocks to increase when the number of different masking phrases used within each block decreases. This kind of increase in interblock variability with decreasing set size was clearly seen in the nonspeech informational masking data of Richards *et al.* (2002). A similar effect is clearly seen in Fig. 5, which shows mean interblock variability for the 14 different conditions tested in this experiment. These interblock variability estimates were calculated from the variability in performance that occurred across the three blocks collected in each of the listening conditions of the experiment. First, the standard deviation was calculated across the three blocks collected in each of the 196 different combinations of listener (7), target talker (2), and masker-randomization condition (14) tested. Then these values were averaged across the 14 combinations of listener and target talker in each randomization condition to generate the mean values shown in Fig. 5. These results confirm that interblock variability was much lower in the nonfrozen conditions (leftmost bar in each group of four bars) than in the three frozen-masker conditions. Within the different frozen-masker conditions, variability was consistently lowest when just content was frozen, and was generally highest when

talker and content were both frozen (white bars in figure). However, despite the fact that the population of masking phrases in the both-frozen blocks of the experiment was much smaller than in the talker-frozen blocks of the experiment (147 phrases versus one), the interblock variability of the both-frozen conditions was only slightly higher in the dichotic listening configurations of the experiment and in fact was slightly lower in the monaural configurations. This suggests that there was substantially more variability in masking effectiveness across the different talkers in the corpus than across the different call-signs, colors, and numbers in the corpus. It should also be noted that overall variability was somewhat lower in the dichotic condition where only the masking talker in the contralateral ear was frozen (third set of bars in Fig. 5). This reflects the overall tendency for variations in the characteristics of the contralateral masker to have relatively little impact on overall performance in the dichotic listening task.

[1]Although the reasons for this decrease are unclear, one possibility is that the listeners in the low-SNR conditions were able to focus their attention on the target talker by listening for the voice that did not change from trial to trial, and that they were unable to use this strategy in configurations where the masking voices were also fixed across the trials in the block.

[2]Although performance was nominally worse when the content was frozen in both ears than when it was only frozen in the target ear (comparing the diamonds in the second and last panels of Fig. 1), the two conditions were not significantly different at any SNR value (Fisher LSD, $p<0.05$). Similarly, although performance was nominally better when both content and talker were frozen in both ears than when both attributes were frozen only in the target ear (comparing the triangles in the second and last panels of Fig. 1), this difference was significant only when the target-ear SNR was -8 dB (Fisher LSD, $p<0.05$).

[3]0 dB was selected because it is the point where informational masking tends to be maximized in the CRM task.

Abouchacra, K., Tran, T., Besing, J., and Koehnke, J. (**1997**). ''Performance on a selective attention task as a function of stimulus presentation mode,'' in Proc. of the Assoc. Res. in Otolaryngology, St. Petersburg, FL.

Bolia, R., Nelson, W., Ericson, M., and Simpson, B. (**2000**). ''A speech corpus for multitalker communications research,'' J. Acoust. Soc. Am. **107**, 1065–1066.

Bronkhorst, A. (**2000**). ''The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions,'' Acustica **86**, 117–128.

Brungart, D. (**2001**). ''Informational and energetic masking effects in the perception of two simultaneous talkers,'' J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D., and Simpson, B. (**2002**). ''Within-channel and across-channel interference in the cocktail-party listening task,'' J. Acoust. Soc. Am. **112**, 2985–2995.

Brungart, D., Simpson, B., Ericson, M., and Scott, K. (**2001**). ''Informational and energetic masking effects in the perception of multiple simultaneous talkers,'' J. Acoust. Soc. Am. **110**, 2527–2538.

Darwin, C., and Hukin, R. (**2000**). ''Effectiveness of spatial cues, prosody, and talker characteristics in selective attention,'' J. Acoust. Soc. Am. **107**, 970–977.

Drullman, R., and Bronkhorst, A. (**2000**). ''Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation,'' J. Acoust. Soc. Am. **107**, 2224–2235.

Egan, J., Carterette, E., and Thwing, E. (**1954**). ''Factors affecting multichannel listening,'' J. Acoust. Soc. Am. **26**, 774–782.

Ericson, M., and McKinley, R. (**1997**). ''The intelligibility of multiple talkers spatially separated in noise,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Hillsdale, NJ), pp. 701–724.

Freyman, R., Helfer, K., McCall, D., and Clifton, R. (**1999**). ''The role of perceived spatial separation in the unmasking of speech,'' J. Acoust. Soc. Am. **106**, 3578–3587.

Hawley, M., Litovsky, R., and Colburn, H. (**1999**). ''Speech intelligibility and localization in a multi-source environment,'' J. Acoust. Soc. Am. **105**, 3436–3448.

Kidd, G. J., Mason, C., and Rohtla, T. (**1995**). ''Binaural advantage for sound pattern identification,'' J. Acoust. Soc. Am. **98**, 1977–1986.

Richards, V., and Tang, Z. (**2002**). ''Cueing Effects in Simultaneous Multi-Tone Informational Masking,'' in Proceedings of the 26th Midwinter Meeting of the Association of Research in Otolaryngology.

Watson, C., Kelly, W., and Wroton, H. (**1976**). ''Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty,'' J. Acoust. Soc. Am. **60**, 1176–1185.

Wright, B., and Saberi, K. (**1999**). ''Strategies used to detect auditory signals in small sets of random maskers,'' J. Acoust. Soc. Am. **105**, 1765–1775.

Development of a new standard laboratory protocol for estimation of the field attenuation of hearing protection devices: Sample size necessary to provide acceptable reproducibility

William J. Murphy[a] and John R. Franks
Hearing Loss Prevention Section, National Institute for Occupational Safety and Health, 4676 Columbia Parkway MS C-27, Cincinnati, Ohio 45226

Elliott H. Berger
E·A·R/Aearo Company, 7911 Zionsville Road, Indianapolis, Indiana 46268-1657

Alberto Behar
University of Toronto, 45 Meadowcliffe Drive, Scarborough, Ontario M1M2X8, Canada

John G. Casali
Virginia Tech, 250 Durham Hall, Blacksburg, Virginia 24061

Christine Dixon-Ernst
Alcoa Corporate Center, 201 Isabella Street, Pittsburgh, Pennsylvania 15212-5828

Edward F. Krieg
Monitoring Research & Statistics Activity, National Institute for Occupational Safety and Health, 4676 Columbia Parkway MS C-22, Cincinnati, Ohio 45226

Ben T. Mozo
Communications & Ear Protection Inc., 303 South Ouida Street, Enterprise, Alabama, 36331

Julia D. Royster and Larry H. Royster
Environmental Noise Consultants, P.O. Box 30698, Raleigh, North Carolina 27622-0698

Stephen D. Simon
Office of Medical Research, Children's Mercy Hospital, 2401 Gillham Road, Kansas City, Missouri 64108

Carol Stephenson
Training and Education Systems Branch, National Institute for Occupational Safety and Health, 4676 Columbia Parkway MS C-10, Cincinnati, Ohio 45226

(Received 30 March 2003; accepted for publication 20 October 2003)

The mandate of ASA Working Group S12/WG11 has been to develop ''laboratory and/or field procedure(s) that yield useful estimates of field performance'' of hearing protection devices (HPDs). A real-ear attenuation at threshold procedure was selected, devised, tested for one earmuff and three earplugs via an interlaboratory study involving five laboratories and 147 subjects, and incorporated into a new standard that was approved in 1997 [Royster *et al.*, ''Development of a new standard laboratory protocol for estimating the field attenuation of hearing protection devices. Part I. Research of Working Group 11, Accredited Standards Committee S12, Noise,'' J. Acoust. Soc. Am. **99**, 1506–1526; ANSI, S12.6-1997, ''American National Standard method for measuring real-ear attenuation of hearing protectors'' (American National Standards Institute, New York, 1997)]. The subject-fit methodology of ANSI S12.6-1997 relies upon listeners who are audiometrically proficient, but inexperienced in the use of HPDs. Whenever a new method is adopted, it is important to know the effects of variability on the power of the measurements. In evaluation of protector noise reduction determined by experimenter-fit, informed-user-fit, and subject-fit methods, interlaboratory reproducibility was found to be best for the subject-fit method. Formulas were derived for determining the minimum detectable difference between attenuation measurements and for determining the number of subjects necessary to achieve a selected level of precision. For a precision of 6 dB, the study found that the minimum number of subjects was 4 for the Bilsom UF-1 earmuff, 10 for the E·A·R Classic earplug, 31 for the Willson EP100 earplug, and 22 for the PlasMed V-51R earplug. [DOI: 10.1121/1.1633559]

PACS numbers: 43.66.Vt, 43.66.Yw [DKW] Pages: 311–323

[a]Electronic address: wmurphy@cdc.gov

I. INTRODUCTION

Royster *et al.* (1996) examined the differences between subject-fit (SF) and informed-user-fit (IUF) hearing protector testing methods. Berger *et al.* (1998) addressed the relationship between the noise reduction measured in the laboratory and the noise reduction measured in occupational settings. The relationship of the variability of real-ear attenuation at threshold (REAT) measurements as a factor controlling sample sizes necessary for adequate statistical power remains an unresolved issue. In this paper, the statistical models for estimating the within-subject repeatability and between-subject and between-laboratory reproducibility are developed and applied to the estimation of the minimum detectable difference and sample-size estimates for REAT measurements.

Hearing protector testing as prescribed by several national and international standards consists of measuring occluded and unoccluded pairs of thresholds in a diffuse sound field with at least ten subjects. The attenuations are measured for at least seven third-octave noise bands (125, 250, 500, 1000, 2000, 4000, and 8000 Hz) as the numerical difference in decibels between occluded and unoccluded threshold pairs calculated for every subject's trial. The difference between the methods studied in this paper deals with the manner in which the subjects and experimenters participated in the fitting of the hearing protection device (HPD).

Royster *et al.* (1996) reported tests on subjects with both SF and IUF methods according to the instructions upon which Method B of ANSI S12.6-1997 (ANSI, 1997) was based. In the SF method, the subjects were provided with only the manufacturer's instructions for wearing the HPD; no experimenter involvement was permitted. They were instructed to fit the device as best they could and then threshold pairs were measured. The IUF method was equivalent to the experimenter-assisted fit in ANSI S12.6-1997 (ANSI, 1997) now commonly referred to as Method A. In this method, the subjects were given the device and the manufacturer's instructions and instructed to fit the HPD. While the subject fitted the device in the presence of fitting noise, the experimenter was permitted to coach the subject to obtain a better fit. As with the SF method, threshold pairs were collected from the subjects. Detailed descriptions of the data collection and subject instructions were given in Royster *et al.* (1996).

An additional study was conducted following the study reported by Royster *et al.* (1996). Subjects were tested at NIOSH (1995) and Virginia Tech University using the subject-fit and the experimenter-fit (EF) methods. The ANSI S3.19-1974 (ANSI, 1974) EF method protocol was used and was equivalent to the test methods presently used by HPD manufacturers to produce data from which noise reduction ratings (NRRs) can be determined. In the EF method, the experimenter fitted the HPD and instructed the subject not to adjust the device during the occluded-threshold tests. From hereon, the testing reported by Royster *et al.* will be referred to as the four-lab study (E·A·RCAL Laboratory, NIOSH Hearing Protector Laboratory, US Army Aeromedical Research Laboratory, Wright Patterson Armstrong Laboratories), while the additional testing performed by NIOSH and Virginia Tech will be referred to as the two-lab study. The results of both studies are reported here.

This paper examines the REAT distributions as a function of protector, fitting procedure, and test frequency. The REATs were analyzed with a multi-level analysis of variance that determined the standard deviations for within-subject repeatability, between-subject reproducibility, and between-laboratory reproducibility. Lastly, the minimum detectable differences at each frequency were determined for each device and fitting procedure. The minimum detectable differences were used to estimate sample sizes necessary to achieve a given desired resolution between measurements.

II. METHODS

A. Data sources

For both the four- and two-lab studies, pairs of occluded and unoccluded thresholds were collected using one-third-octave band noise stimuli centered at 125, 250, 500, 1000, 2000, 4000, and 8000 Hz. The hearing protectors tested, the test methods, and studies providing the data reported here are summarized in Table I.

1. Four-lab study

The HPDs that were selected for the four-lab study are described in Sec. II C and Fig. 1 of Royster *et al.* (1996). They included the E·A·R® Classic® foam earplug, the PlasMed, Inc. V-51R premolded earplug (five sizes), the Willson Safety Products EP100 premolded earplug (two sizes), and the Bilsom UF-1 earmuff. The devices were selected based upon availability of published real-world data, market-place popularity at the commencement of the studies, and diversity of protector types. The Working Group had selected more earplugs than earmuffs because it determined that earplugs provided a greater real-world estimation problem than did earmuffs (Casali and Park, 1991).

Subjects for the four- and two-lab experiments were selected from volunteers with pure-tone air-conducted hearing thresholds less than 25 dB HL *re* ANSI S3.6-1999 (ANSI, 1999) at each test frequency and normal tympanometry. Af-

TABLE I. Hearing protection devices, numbers of subjects, and methods of fit used in the four- and two-lab studies.

Devices	Methods	
	Four-lab study	Two-lab study
	E·A·RCAL (26 subjects) USAARL (24 subjects) NIOSH (24 subjects) WPAFB (24 subjects) (Royster *et al.*, 1996)	NIOSH (25 subjects) Virginia Tech (26 subjects) (Franks *et al.*, 2000)
Bilsom UF-1 earmuff	SF[a] IUF[b]	
Willson EP100 earplug	SF[a] IUF[b]	
E·A·R Classic earplug	SF[a] IUF[b]	SF[a] EF[c]
PlasMed V-51R earplug	SF[a] IUF[b]	SF[a] EF[c]

[a]SF: Subject fit as defined in ANSI S12.6-1997, Method B (ANSI, 1997).
[b]IUF: Informed-user fit equivalent to experimenter assisted fit as defined in ANSI S12.6-1997, Method A, (ANSI, 1997).
[c]EF: Experimenter fit as defined in ANSI S3.19-1974 (ANSI, 1974).

ter an initial session for hearing screening and audiometric training, each subject made eight visits to the lab. In each visit, two different HPDs were tested in separate sessions; each included two trials pairing sets of unoccluded and occluded sound-field thresholds. In total, four unoccluded and occluded paired thresholds were measured for each device, subject, and fitting method. Subjects received a brief rest break between the two sessions of a visit. Visits were separated by a minimum of 6 h, and all eight visits were required to take place within 21 days. Each laboratory recruited 24 subjects who were naive in the use and fitting of hearing protectors. The subject population was gender-balanced and the order of device testing was randomized to minimize learning effects that might occur. REAT means and standard deviations for the four-lab study were reported in Table II of Royster *et al.* (1996). The instructions regarding fitting and size selection are given in detail by Royster *et al.* (1996). For the sized earplugs, subjects were allowed to select the size that best fit their ear canals. The experimenter provided no assistance during the subject-fit portion of the testing and provided limited coaching during the informed-user fit testing.

2. Two-lab study

The two-lab study tested the V-51R and the E·A·R Plug using SF and EF methods. The NIOSH and Virginia Tech laboratories tested REAT for 25 and 26 subjects, respectively, first with the SF method and then with the EF method. The test methods and data have been reported in detail elsewhere (Franks *et al.*, 2000). After an initial session for hearing screening and audiometric training, each subject visited the lab on separate days for SF and EF testing. In each visit, three pairs of unoccluded and occluded sound-field thresholds were collected for two different HPDs in separate sessions. Subjects were permitted a brief rest break as needed between paired trials during a visit. These data were collected in order to obtain a direct comparison of the REATs measured from the two methods on the same subjects, rather than relying solely on the SF data from the four-lab study and upon the manufacturer's report of EF data.

B. Data analyses

The analyses of the data were conducted in three phases. First, the distributions of the REAT histograms for each fitting method, device, and test frequency were tested for normality and modality. Second, the REAT distributions were analyzed with a multi-level analysis of variance (ANOVA) with trials nested within subjects and subjects within laboratories. Third, the number of subjects necessary to achieve a given resolution in the REAT data were calculated for a confidence level $1-\alpha=0.84$ and a power of $\beta=0.80$. The confidence level represents one standard deviation away from the mean. The confidence level is one minus the probability of committing a type I error, incorrectly identifying an effect. The power is one minus the probability of committing a type II error, failure to identify a real effect.

The statistical design accounts for three sources of variation: trial-to-trial (σ_{trial}), subject-to-subject (σ_{subject}), and laboratory-to-laboratory ($\sigma_{\text{laboratory}}$). From these sources, estimates for the within-subject repeatability, between-subject reproducibility, and interlaboratory reproducibility were calculated. The data from each fitting method (EF, IUF, and SF) were analyzed separately. The EF data represent the two-lab study; the IUF data represent the four-lab study and the SF data represent the pooled results from four- and two-lab studies. Learning effects due to repeated tests with the same protector (within-subject repeatability), subject effects for a given hearing protector (between-subject repeatability), or laboratory effects (between subject/between laboratory repeatability) were assessed by analysis of variance. The geographical separation of laboratories prevented using the same test subjects for directly assessing interlaboratory effects.

III. RESULTS

A. Distribution of REATs

The data for each HPD were pooled across the labs and examined by frequency and fitting method. The REAT means and standard deviations from the four-lab study have been reported for the SF and IUF methods for all four devices in Royster *et al.* (1996). Royster *et al.* did not examine the normality and modality of the REAT distributions. While the results of the two-lab study were not substantially different from the four-lab study, the additional data for SF and EF methods provided the impetus for this analysis. The REAT distributions for the pooled data from both studies are shown in Figs. 1–3. The figures depict at each test frequency histograms of the attenuation data sorted into 3-dB-wide bins. For example, if an attenuation estimate for a subject was greater than or equal to −1.5 dB and less than 1.5 dB, it was counted as one occurrence in the 0-dB bin. Similarly, the other bins were centered on multiples of 3 dB. The width of each bar represents the proportional number of occurrences at each attenuation level. In the upper left panel of Figs. 1–3, a scale bar the width of 50 occurrences is shown. The color represents a transition from minimum to maximum (blue to red), where the red bar indicates the maximum of the distribution for a given test frequency. Bins with a value of zero are not given a color or width. The diamond symbol represents the mean attenuation for a given test frequency.

Each distribution was tested for normality with the SAS univariate procedure (SAS, 1998) and was compared to three different probability distributions: normal, gumbel, and mixed-normal. In an extension to the maximum likelihood analysis in Murphy *et al.* (2002), the distribution was determined to be better estimated by one probability distribution or another. For instance, a distribution might be more likely to have been sampled from a gumbel distribution than from a normal distribution. The probability estimates of the maximum likelihood procedures were used according to Table II to classify whether a distribution was normal, gumbel, mixed-normal, or non-normal. If the SAS univariate procedure showed that the distribution was significantly different from a normal distribution, then the remainder of the tests were used to determine the class. In Figs. 1–3, the classifications of the distributions (non-normal, gumbel, and mixed-normal) are shown for each protector at each test frequency

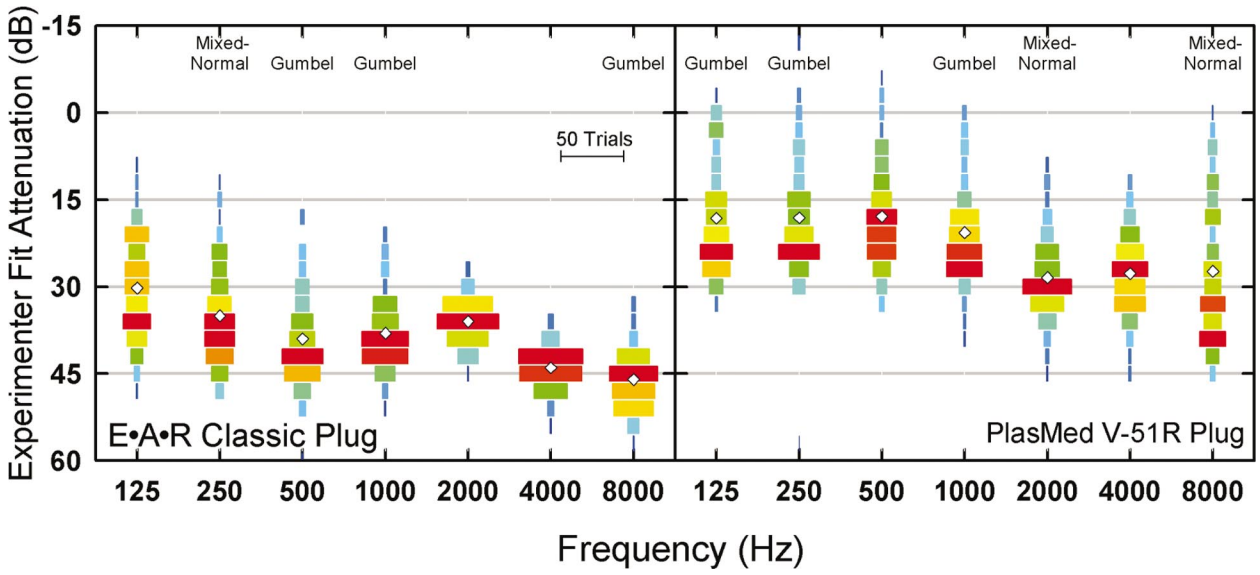


FIG. 1. Histogram-frequency plot of the attenuation measurements for the experimenter-fit method. The EF method data were collected as part of the two-lab study and represent 51 subjects with three occluded–unoccluded trials per subject per protector. The width of the bars represent the number of trials where subjects achieved an attenuation within a given 3-dB bin. A scale of 50 trials is shown in the upper left panel. The colored bars indicate the maximum (red) and minimum (blue) number of occurrences of a particular attenuation within a given frequency band. The colors were independently scaled for each frequency band distribution. The diamond symbols denote the mean of the distribution. The shape of each distribution has been tested and the classification according to Table II are listed above each histogram: ''Gumbel,'' ''Mixed-normal,'' and ''Non-normal.'' Histograms without a label are not significantly different from normal.

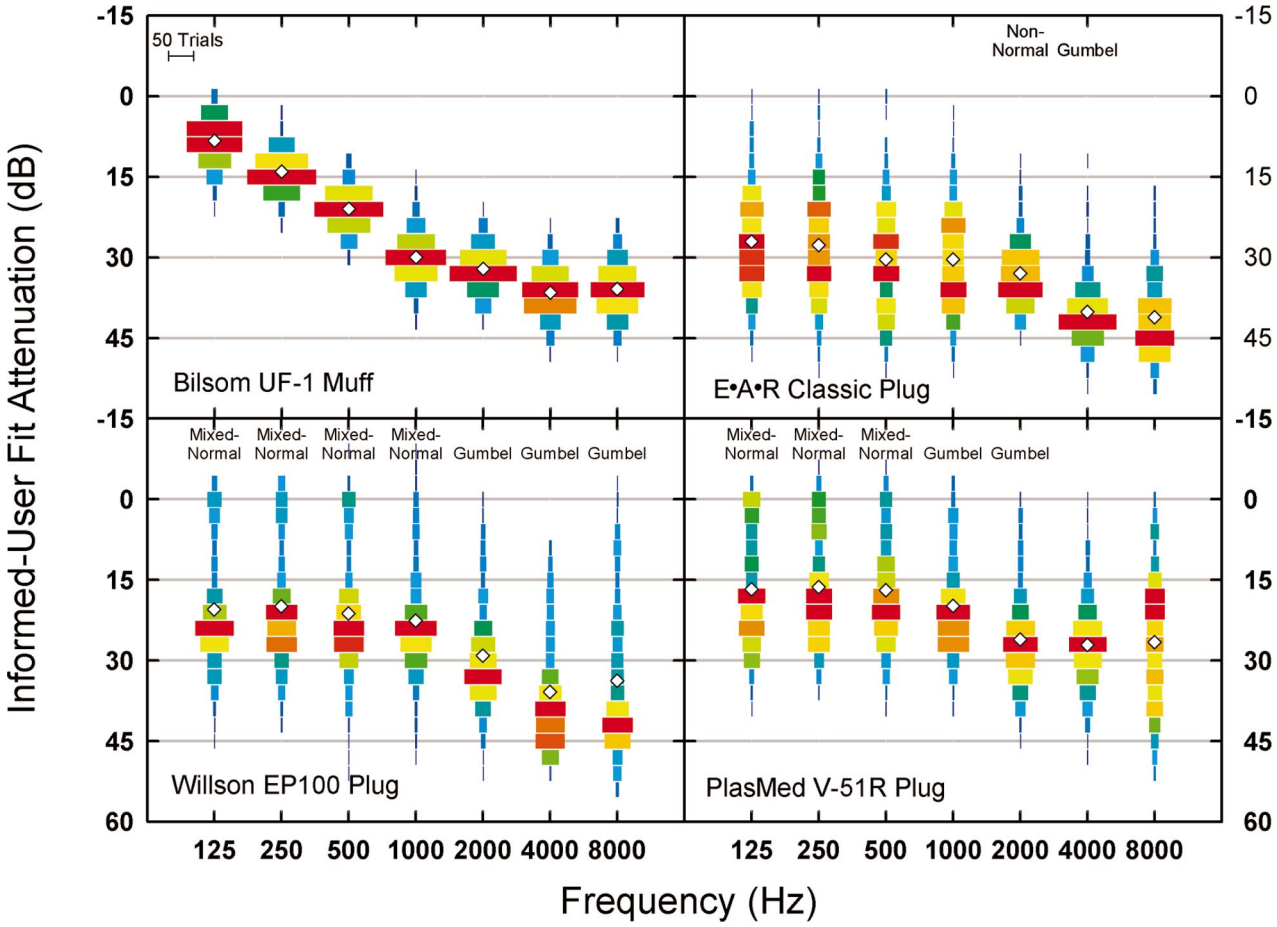


FIG. 2. Histogram-frequency plot of the attenuation measurements for the informed-user fit method. The IUF method data were collected as part of the four-lab study and represent 96 subjects with four occluded-unoccluded trials per subject per protector.

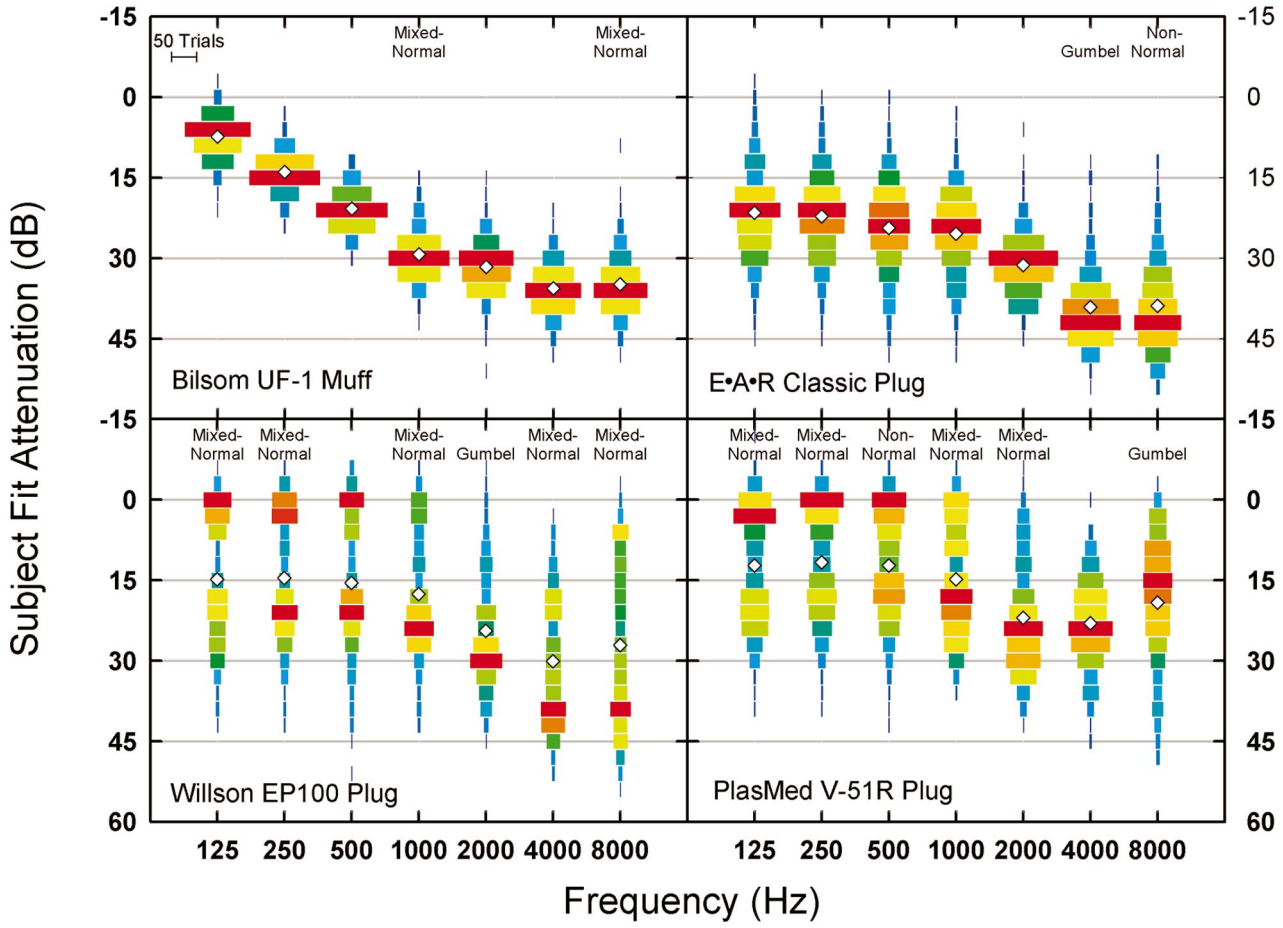


FIG. 3. Histogram-frequency plot of the attenuation measurements for the subject-fit method. The SF method data were collected as part of the four- and two-lab studies and represent 147 subjects with four occluded-unoccluded trials per subject per protector in the four-lab study and three occluded-unoccluded trials per protector in the two-lab study.

above each distribution. If a distribution was not significantly different from normal, there is no classification label. Although the figures do not display individual laboratory data, normal, gumbel, mixed-normal, and non-normal distributions are evident in the individual laboratory data as well as the pooled results of all labs.

The two-lab EF data for the E·A·R plug and V-51R earplug are shown in Fig. 1. The REAT distribution for 250 Hz was judged to be mixed-normal for the E·A·R plug. The REAT distributions for 500, 1000, and 8000 Hz were judged to be best fit by the gumbel distribution for the E·A·R plug. The EF data for the V-51R earplug were normal at 500 and 4000 Hz. The mixed-normal model yielded a better fit at 2000 and 8000 Hz. The distributions for 125, 250, and 1000 Hz were best fit by the gumbel distribution.

In Fig. 2, four-lab IUF data for the Bilsom UF-1 earmuff exhibited normal distributions for all frequencies. The IUF REAT distributions for the E·A·R plug were non-normal for 2000 Hz and best fit by the gumbel distribution for 4000 Hz. The REAT distributions for the EP100 earplug were mixed-normal for 125, 250, 500, and 1000 Hz. The distributions at 2000, 4000, and 8000 Hz were best fit with the gumbel model. The distributions at lower frequencies exhibited a clear tendency of one mode centered around 0 dB while at higher frequencies the distributions are skewed. The V-51R REAT distributions were best fit with the gumbel model at

TABLE II. Rules for classification of REAT distributions in Figs. 1–3 for a significance $p<0.05$. For instance, the probabilites for the EP100 earplug SF REAT distribution at 2000 Hz were $p_{SAS}=0.0042$, $p_{MLEGumbel\text{-}Normal}=0.015$, $p_{MLEMixedNormal\text{-}Normal}=0.0044$, and $p_{MixedNormal\text{-}Gumbel}=0.19680$. According to the rules, this distribution was classified as Gumbel.

Classification	SAS univariate normality	Maximum likelihood gumbel-normal	Maximum likelihood mixed normal-normal	Maximum likelihood mixed normal-gumbel
Normal	Not significant	Not applicable	Not applicable	Not applicable
Mixed-normal	Significant	Significant	Significant	Significant
Gumbel	Significant	Significant	Significant	Not significant
Mixed-normal	Significant	Not significant	Significant	Not applicable
Gumbel	Significant	Significant	Not significant	Not applicable
Non-normal	Significant	Not significant	Not significant	Not applicable

TABLE III. Standard deviations for the trial effects, σ_{trial}, for each device, fit, and frequency in decibels (dB).

		Frequency (Hz)						
Device	Fit	125	250	500	1000	2000	4000	8000
Bilsom	IUF	2.5	2.5	2.7	2.6	2.5	2.5	3.1
UF-1 earmuff	SF	2.2	2.5	2.7	2.5	2.6	2.7	3.1
EAR	EF	5.1	4.5	4.5	3.5	2.4	2.4	2.5
Classic earplug	IUF	4.9	4.9	4.5	4.2	3.1	3.3	3.9
	SF	5.7	5.5	5.4	4.7	3.1	3.7	4.5
Willson	IUF	5.8	6.0	5.8	5.2	4.8	5.0	6.5
EP100 earplug	SF	6.6	6.7	6.9	5.7	5.2	5.9	7.9
PlasMed	EF	5.6	5.1	4.5	4.4	4.0	4.2	5.6
V-51R earplug	IUF	6.2	5.9	6.0	5.5	4.5	5.1	6.8
	SF	5.7	5.5	5.2	5.1	4.6	3.9	5.4

1000 and 2000 Hz. The mixed-normal model provided a better fit at 125, 250, and 500 Hz. Similar trends of low attenuation at 125, 250, and 500 Hz were evident for the V-51R REAT distributions as are evident in the EP100 data.

In Fig. 3, the combined two- and four-lab SF REAT distributions for the Bilsom UF-1 muff were mixed-normal at 1000 and 8000 Hz. The REAT distribution for the E·A·R plug was best fit with the gumbel model at 4000 Hz and at 8000 Hz was non-normal. For the EP100 earplug, the REAT distributions were mixed-normal at 125, 250, 1000, 4000, and 8000 Hz. At 2000 Hz the distribution was best fit with the gumbel model. The EP100 REAT distribution at 500 Hz was not significantly different from a normal distribution, but it has the features of a bimodal distribution. The V-51R REAT distributions were non-normal at 500 Hz, mixed-normal at 125, 250, 1000, and 2000 Hz, and gumbel at 8000 Hz. The distribution at 4000 Hz was normally distributed. The low attenuation modes were apparent for the EP100 and V-51R earplugs.

B. Standard deviations

A multi-level analysis of variance (Netter, 1990) was used to estimate the standard deviations for laboratory, subject, and trial effects. The statistical model was

$$Y_{ijk} = \mu + \text{Trial}_{k(ij)} + \text{Subject}_{j(i)} + \text{Lab}_i, \quad (1)$$

where Y_{ijk} is the measured attenuation, μ is the real attenuation, and $\text{Trial}_{k(ij)}$ is the random error term for the kth trial within the jth subject and ith laboratory, $\text{Subject}_{j(i)}$ is the random error term for the jth subject within the ith laboratory, and Lab_i is the random error term for the ith laboratory. The ANOVA calculations were performed using the S-Plus software package (S-Plus, 2002). The model results are presented in Tables III–V.

Table III presents the standard deviations in decibels for the trial effects, σ_{trial}, determined from both the four- and two-laboratory studies for each of the fitting methods. Smaller standard deviations indicate less variability in the effect of the experimenter or subject on fitting the hearing protector consistently. Comparing the four sets of IUF and SF data, the Bilsom earmuff exhibited the smallest standard deviations, which is not surprising since earmuffs are easier to fit and offer fewer opportunities for improper fitting. The E·A·R plug tended to have the next smallest standard deviations. The EP100 and V-51R devices tended to have larger standard deviations. In general, as experimenter involvement decreased, σ_{trial} increased.

Table IV presents the standard deviations for the subject effects, σ_{subject}. Again, comparing the four sets of IUF and SF data, the Bilsom earmuff exhibited the smallest standard deviations. The E·A·R plug had the next smallest deviations followed by the V-51R and EP100, respectively. The small

TABLE IV. Standard deviations for the subject effects, σ_{subjects}, for each device, fit, and frequency in decibels (dB).

		Frequency (Hz)						
Device	Fit	125	250	500	1000	2000	4000	8000
Bilsom	IUF	2.9	2.0	1.9	2.9	2.4	2.8	2.9
UF-1 earmuff	SF	2.8	2.2	2.0	2.8	3.2	2.9	3.8
E·A·R	EF	5.6	6.2	5.6	4.7	2.6	2.0	3.8
Classic earplug	IUF	5.6	5.8	5.9	5.0	3.3	3.5	3.9
	SF	5.5	5.1	6.1	5.1	3.6	4.4	5.7
Willson	IUF	8.5	8.2	8.6	7.4	6.9	7.9	9.9
EP100 earplug	SF	9.6	9.4	10.4	9.4	8.7	9.4	11.5
PlasMed	EF	6.7	6.8	6.8	6.3	5.2	5.0	9.9
V-51R earplug	IUF	7.1	6.9	6.8	6.4	6.7	5.2	9.1
	SF	8.5	8.2	8.3	8.5	8.5	6.6	10.1

TABLE V. Standard deviations for the laboratory effects, $\sigma_{\text{laboratory}}$, for each device, fit, and frequency in decibels (dB). Standard deviations less than 0.05 were rounded to 0.0 dB.

		Frequency (Hz)						
Device	Fit	125	250	500	1000	2000	4000	8000
Bilsom	IUF	0.8	0.3	0.8	1.6	1.5	1.3	1.1
UF-1 earmuff	SF	0.0	0.0	0.5	1.8	1.5	1.4	0.7
E·A·R	EF	2.1	0.9	0.0	0.0	0.4	0.0	0.0
Classic earplug	IUF	4.4	5.4	5.8	5.6	2.7	2.4	3.8
	SF	1.0	1.3	1.6	2.3	2.0	2.4	2.7
Willson	IUF	2.9	2.6	3.2	2.7	3.3	3.0	4.5
EP100 earplug	SF	1.0	0.0	0.0	0.0	2.1	1.5	2.2
PlasMed	EF	0.0	0.0	0.0	0.0	1.2	0.0	0.0
V-51R earplug	IUF	2.9	3.0	2.7	1.7	2.2	2.4	4.4
	SF	0.8	1.2	1.0	1.2	1.9	1.8	0.0

variance for the Bilsom earmuff is indicative of the consistency of the fit across subjects. As with σ_{trial}, σ_{subject} increased with decreasing experimenter intervention.

Table V presents the standard deviations for the laboratory effects, $\sigma_{\text{laboratory}}$. As can be seen, the standard deviations for the SF method, where there was minimal experimenter involvement, were almost always lower than the values for the IUF method where the experimenter provided advice. However, $\sigma_{\text{laboratory}}$ for the EF method, where the experimenter was fully involved, were generally lower. For several frequencies and fitting methods, the standard deviations were less than 0.05 dB and were rounded to 0.0 dB.

C. Repeatability and reproducibility

Within-subject repeatability is a measure of the consistency of attenuation across trials for the same sample of subjects and hearing protector. Within-subject repeatability is computed with σ_{trial} listed in Table III. Between-subject reproducibility is a measure of the consistency of attenuation across subjects and trials for a hearing protector. Between-subject reproducibility is computed with the σ_{trial} and σ_{subject} from Tables III and IV. If there is no change of the testing protocol over time, then between-subject reproducibility measures consistency when two hearing protectors are tested with different subject panels. Between-laboratory reproducibility incorporates the lab-to-lab standard deviation, $\sigma_{\text{laboratory}}$, and measures the consistency of attenuations across laboratories. If there is a negligible amount of lab-to-lab variation, then between-subject reproducibility measures the consistency between two laboratories. While subjects were randomly sampled, the laboratories were not randomly selected as would be necessary to give a true estimate of laboratory effects.

Within-subject repeatability was estimated with the equation,

$$\sigma_{\text{within-subject}} = \sqrt{\frac{\sigma_{\text{trial}}^2}{(n_s n_t)}}, \tag{2}$$

where σ_{trial}^2 was the trial-to-trial variance, n_s was the number of subjects, and n_t was the number of trials per subject.

Between-subject reproducibility was estimated with the equation

$$\sigma_{\text{between-subject}} = \sqrt{\frac{\sigma_{\text{subject}}^2}{n_s} + \frac{\sigma_{\text{trial}}^2}{(n_s n_t)}}, \tag{3}$$

where $\sigma_{\text{subject}}^2$ was the subject-to-subject variance.

When $\sigma_{\text{laboratory}}$ is zero or negligible, then reproducibility $\sigma_{\text{between-subject}}$ represents consistency between laboratories. When $\sigma_{\text{laboratory}}$ is large, then comparisons of hearing protectors between laboratories are inappropriate. Between-laboratory reproducibility was estimated with the equation

$$\sigma_{\text{between-laboratory}} = \sqrt{\sigma_{\text{laboratory}}^2 + \frac{\sigma_{\text{subject}}^2}{n_s} + \frac{\sigma_{\text{trial}}^2}{(n_s n_t)}}, \tag{4}$$

where $\sigma_{\text{laboratory}}^2$ was the laboratory-to-laboratory variance. The standard deviation estimates of $\sigma_{\text{within-subject}}$, $\sigma_{\text{between-subject}}$, and $\sigma_{\text{between-laboratory}}$ calculated from Tables III–V for each device and each test frequency are shown in Figs. 4–6, respectively, using $n_s = 20$ and $n_t = 2$. These are similar to the repeatability and reproducibility concepts recommended for use in interlaboratory studies by ISO 5725-2 (ISO, 1994). Results for $n_t = 3$ or 4 are not shown, but can easily be calculated from these formulas.

1. Within-subject repeatability

The standard deviations for within-subject repeatability, $\sigma_{\text{within-subject}}$, for the Bilsom UF-1 earmuff in Fig. 4 exhibit little difference between SF and IUF methods and across test frequencies.

The standard deviations for all fit protocols are greater for the E·A·R plug than for the earmuff except at 4000 and 8000 Hz, EF fit. The standard deviations are dependent upon test frequency, being the lowest at 2000 Hz. As well, the magnitudes of the standard deviations increase above 2000 Hz as the amount of experimenter intervention decreases; the lowest standard deviations can be observed for the EF method while the highest standard deviations are for the SF method.

The standard deviations for the EP100 earplug are the highest of the four devices tested during these investigations. The standard deviations for the SF method are generally 0.1

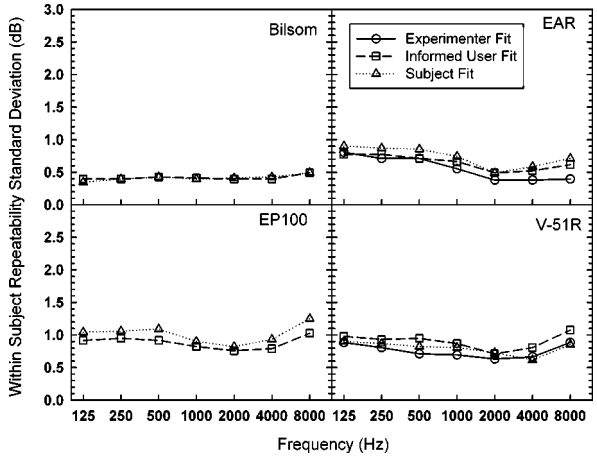


FIG. 4. The standard deviations for within-subject repeatability for each protector by fitting method and frequency. The REAT data was analyzed with a multi-level ANOVA where within-subject effects were nested within subjects and between-subject effects were nested within laboratory effects. The Bilsom earmuffs had the lowest variance and exhibited little difference between fitting methods. The other devices exhibited more variance and a trend to increased variability with decreased experimenter involvement. Experimenter-fit data are from the two-lab study, while subject-fit data are from the four-lab study.

to 0.2 dB greater than the standard deviations for the IUF method. Again, the magnitudes of the SF standard deviations are frequency dependent, being lowest at 2000 Hz.

The V-51R earplug exhibits higher standard deviations than the E·A·R plug, but lower than the EP100 earplug. As with the other earplugs, standard deviations tended to be lowest at 2000 Hz. However, for this earplug, the SF values were somewhat lower than the IUF values except at 2000 Hz where they were equal. The EF method gave the lowest standard deviations where the EF and SF methods values were virtually identical.

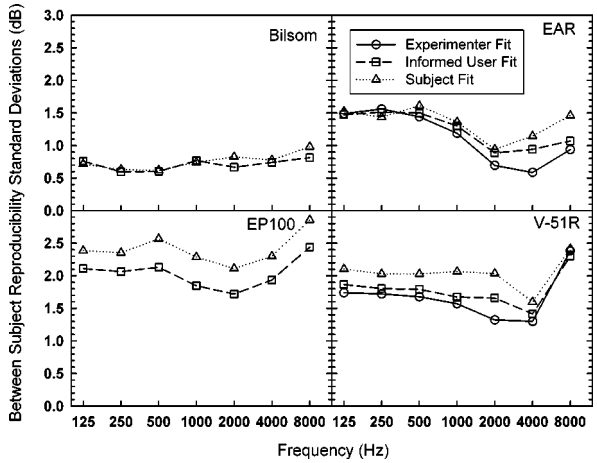


FIG. 5. The standard deviations for between-subject reproducibility for each protector by fitting method and frequency. The REAT data was analyzed with a multi-level ANOVA where within-subject effects were nested within subjects and between-subject effects were nested within laboratory effects. The Bilsom earmuffs exhibited the smallest standard deviations with almost identical results between IUF and SF fitting methods. The earplugs exhibited comparable standard deviations across devices with a slight trend for increased standard deviations with decreasing experimenter involvement. Experimenter-fit data are from the two-lab study, while subject-fit data are from the four-lab study.

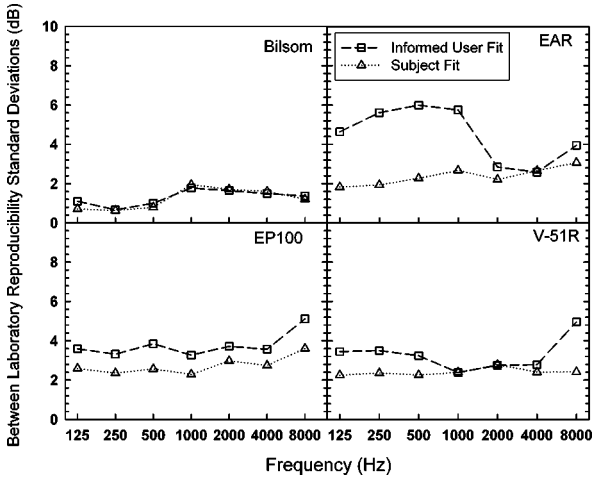


FIG. 6. The standard deviations for between-laboratory reproducibility. The magnitude of the standard deviations, greater than 2 dB, for the earplugs indicate poor reproducibility. The data from the Bilsom earmuffs indicate highly reproducible results across labs. The E·A·R Plug exhibits greater standard deviations for the IUF method. The standard deviations for the IUF method are greatest for almost all test frequencies for every device.

The results for the within-subject repeatability show consistent performance across test methods. The repeatability standard deviations depend upon the type of protector. Although slight differences exist between the standard deviations as a function of experimenter involvement, the differences are too small to be considered meaningful.

2. Between-subject reproducibility

The standard deviations for between-subject reproducibility, $\sigma_{\text{between-subject}}$, for each of the tested hearing protectors are shown in Fig. 5. In general the lowest standard deviations are for the Bilsom UF-1 earmuff and the highest are for the EP100 premolded earplug. The SF standard deviations are approximately equal to the IUF standard deviations for the Bilsom UF-1 earmuff. For the EP100 premolded earplug, and the V-51R premolded earplug, the IUF standard deviations for reproducibility are uniformly less than the SF standard deviations. For the E·A·R plug, the IUF standard deviations are also less than the SF standard deviations except at 250 Hz. For the Bilsom UF-1, differences between IUF and SF are trivial. However, these differences are not trivial for the E·A·R plug at 4000 and 8000 Hz, for the EP100 at all frequencies, and for the V-51R at all frequencies except 4000 and 8000 Hz.

In the two-lab study, the SF standard deviations are greater than the EF standard deviations for the V-51R earplug below 8000 Hz. For the E·A·R plug, the SF standard deviations are greater than the EF standard deviations except at 250 Hz.

3. Between-laboratory reproducibility

The standard deviations for between-laboratory reproducibility, $\sigma_{\text{between-laboratory}}$ are shown in Fig. 6. The smallest deviations are those of the Bilsom UF-1 earmuff and the largest are for the E·A·R plug IUF data. In contrast to the between-subject reproducibility, the SF standard deviations

are less than or approximately equal to the IUF standard deviations for nearly every frequency and protector.

D. Minimum detectable differences and sample size estimates

For a method to be useful in providing numbers for rating systems, the predictiveness of the data must be determined. In general, as sample size increases, so does the precision of the predictions of protector performance. In these cases, when the desired precision for a protector is established, the number of subjects necessary to achieve a given resolving ability can be estimated for a given confidence level, $1-\alpha$, and power, $1-\beta$ (e.g., confidence level of 84%, $\alpha=0.16$; power of 80%, $\beta=0.20$). The predictive precision, or resolution, of a hearing protector may be expressed as the smallest difference in two attenuation measurements that can be reliably detected for a given confidence level and a power level.

The confidence level was selected based upon the approach used in determining the noise reduction rating for subject fit data (NRR_{SF}); the calculation subtracts one standard deviation from the mean attenuation which yields an minimum attenuation estimate for 84% of the population at a given frequency (Franks *et al.*, 2000). The power level of 0.80 typically selected for behavioral data such as hearing thresholds assumes normality of the data.

1. Minimum detectable differences

Power calculations for testing both the within-subject repeatability and between-subjects reproducibility involve a simple modification of the formulas for power for a two-sample *t*-test (Rosner, 1990). For a hypothesis without multiple sources of variation and assuming the data are drawn from a normal distribution, the minimum detectable difference, D, with power, $1-\beta$ and confidence level, $1-\alpha$, is

$$D=(\text{Probit}(1-\alpha)+\text{Probit}(1-\beta))\frac{\sqrt{2}\sigma}{\sqrt{n}}, \tag{5}$$

where the Probit function provides the appropriate percentile value from a standard normal distribution (SAS, 1998). The minimum detectable difference is the value below which differences in real-ear attenuations at threshold are statistically insignificant.

Notice that the final term in the above equation, $\sqrt{2}\sigma/\sqrt{n}$, is simply the standard error of the difference between the means. Assuming that the between-subject reproducibility hypothesis to be tested is based on n_s subjects with n_t trials apiece, the minimum detectable difference becomes

$$D=(\text{Probit}(1-\alpha)+\text{Probit}(1-\beta))\sqrt{2}\sigma_{\text{between-subject}}, \tag{6}$$

where $\sigma_{\text{between-subject}}$ from Eq. (3) replaces $\sigma/\sqrt{n}$ in Eq. (5). As expected, increasing the number of subjects, n_s, decreases the contributions of σ_{subject} and σ_{trial} to the standard error term. Quite often, the minimum detectable difference will be smaller than the meaningful difference and it is always more influenced by the number of test subjects than the number of repetitions of test conditions (trials) per subject. Increasing the number of trials, n_t, decreases only the contribution of σ_{trial} to the standard error. The between-subject reproducibility test is valid when the between-laboratory reproducibility is small. As a rule of thumb, $\sigma_{\text{between-laboratory}}$ should be no more than 10% of the minimum detectable difference. The effect of subject sample size and number of trials is illustrated in Fig. 7. The σ_{trial} and σ_{subject} were assumed to be 5 dB. Doubling the number of subjects improves the minimum detectable difference more than doubling or tripling the number of tests per subject.

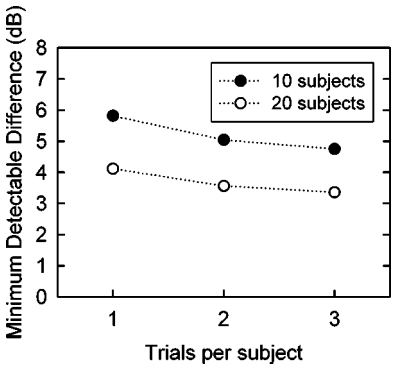


FIG. 7. Effects of number of subjects and number of trials per subject on the minimum detectable difference. The example shown is for the Bilsom UF-1 earmuff. The same relationship applies to the other devices tested—doubling the number of subjects is more effective than doubling the number of trials per subject. The parameters used to calculate the data points were $\sigma^2_{\text{trial}}=5$ dB, $\sigma^2_{\text{subject}}=5$, $n_s=10$, 20, and $n_t=1$, 2, 3.

Since the within-subject repeatability test uses the same subjects for the hearing protectors under test, the between-subject source of variation that would be nested within protectors is eliminated. For this design, the minimum detectable difference may be computed as

$$D=(\text{Probit}(1-\alpha)+\text{Probit}(1-\beta))\sqrt{2}\sigma_{\text{within-subject}}. \tag{7}$$

The within-subject repeatability test is valid so long as the same subjects are used for testing both hearing protectors. If different subjects are used, as might be required if there is a time gap between the testing of the two hearing protectors, then one should use the between-subject reproducibility hypothesis. When using this hypothesis, a laboratory needs to demonstrate that there are no significant trends over time relative to the size of the difference they are trying to detect. Table VI shows the minimum detectable differences calculated using $1-\alpha=0.84$, $1-\beta=0.80$, $n_s=20$ and $n_t=2$, and the between-subject reproducibility found in Table IV. For the Bilsom earmuff, the minimum detectable differences ranged between 1.6 at 250 and 500 Hz and 2.6 dB at 8000 Hz for the SF method. For the E·A·R plug, the SF method yielded differences ranging from 2.5 to 4.2 dB. For the EP100 and V-51R, the ranges of the minimum detectable differences for the SF method were 5.5 to 7.4 dB and 4.2 to 6.3 dB, respectively. Small minimum detectable differences for the Bilsom and EAR protectors indicate that a given protector is more likely to be uniformly fit across subjects.

TABLE VI. Minimum detectable differences for each device, fit, and frequency in dB. Note: $1-\alpha=0.84$, $1-\beta=0.8$, $n_s=20$, $n_t=2$. The between-subject standard deviations are taken from Table IV.

		Frequency (Hz)						
Device	Fit	125	250	500	1000	2000	4000	8000
Bilsom	IUF	2.0	1.6	1.6	2.0	1.7	1.9	2.1
UF-1 earmuff	SF	1.9	1.6	1.6	1.9	2.1	2.0	2.6
E·A·R	EF	3.9	4.1	3.7	3.1	1.8	1.5	2.4
Classic earplug	IUF	3.8	3.9	3.9	3.4	2.3	2.4	2.8
	SF	4.0	3.7	4.2	3.5	2.5	3.0	3.8
Willson	IUF	5.5	5.4	5.5	4.8	4.5	5.0	6.4
EP100 earplug	SF	6.2	6.1	6.7	6.0	5.5	6.0	7.4
PlasMed	EF	4.5	4.5	4.4	4.1	3.4	3.4	6.2
V-51R earplug	IUF	4.9	4.7	4.7	4.4	4.3	3.7	6.0
	SF	5.5	5.3	5.3	5.4	5.3	4.2	6.3

2. Number of subjects necessary for a desired resolution

The equation for calculating precision based on the variance for the sample size tested is

$$N_{\text{subjects}}=n_s\left(\frac{D}{R}\right)^2, \tag{8}$$

where N_{subjects} is the estimated sample size, n_s is the sample size for the tested population, D is the minimum detectable difference determined from the tested population for a given power and confidence level, and R is the target resolution in decibels. Minimum detectable difference is equivalent to the desired resolution except that the desired resolution is chosen rather than determined from the tested population. This formula has been applied to the minimum detectable differences derived from the four- and two-lab studies for a resolution of 6 dB with an $\alpha=0.16$ and $\beta=0.20$ for each protector and fitting condition. The desired resolution is the figure of merit of the ability to distinguish between two distributions of attenuations at any of the test frequencies. If, for instance, two sets of attenuations had been measured from earmuffs from separate production runs, the resolution and the minimal detectable difference could be used to determine how many subjects need to be tested to identify any manufacturing differences in the devices.

Figure 8 displays the estimated number of subjects, N_{subjects}, calculated from the minimum detectable difference in Eq. (6) for the three fitting methods using $\sigma_{\text{between-subject}}$ with $n_s=20$ subjects, $n_t=2$ trials, and $R=6$ dB. These values were selected according to the sample sizes, repetitions, and desired resolution used in the ANSI S12.6-1997 (ANSI, 1997) standard. While results for a given protector will not yield the same results presented here, these estimates represent conservative estimates from a variety of protectors. The preceding calculations should be performed for a particular data set to determine whether the minimum number of subjects has been achieved to reach the 6-dB minimum detectable difference.

The Bilsom UF-1 earmuff exhibited the smallest estimated sample sizes—4 subjects for the SF method at 8000 Hz. Likewise, the E·A·R plug exhibited a somewhat larger sample size than the earmuff (10 subjects for the SF method at 500 Hz), but considerably smaller than the two premolded earplugs. The EP100 earplug required 31 subjects for a resolution of 6 dB (8000 Hz), while the V-51R earplug needed 22 subjects for a 6-dB (the distance between distributions) resolution (also driven by 8000 Hz). The number of subjects must be rounded up to the nearest integer (e.g., 3.4 would round up to 4 subjects).

IV. DISCUSSION

A. REAT histograms

The REAT data presented in Figs. 1–3 illustrate the effect of fitting method upon the quality of fit. For an earmuff such as the Bilsom UF-1, the fit can be affected by disruption of the seal of the cushion such as by the ear pieces of safety glasses. Discounting damage to the cushions and improper placement of the muff over the pinna, earmuffs are easily fit on a subject's head. The agreement between IUF and SF REAT histograms demonstrate that additional experimenter involvement has little effect on the attenuation results for

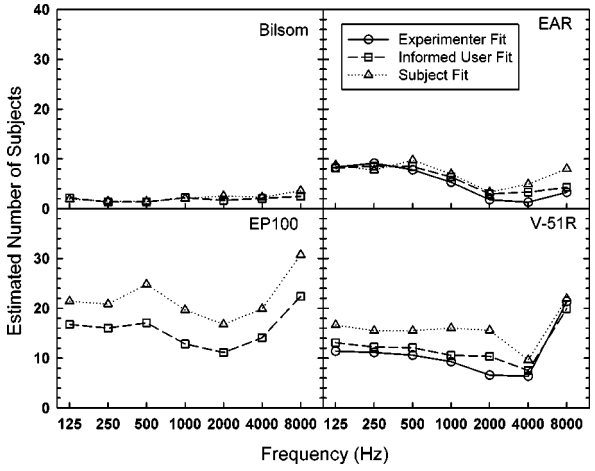


FIG. 8. Sample sizes necessary to achieve a 6-dB resolution in minimum detectable difference in attenuation determined by a power calculation based upon the repeatability and reproducibility analysis. The power calculations were based on $\alpha=0.16$ and $\beta=0.20$ which is equivalent to an NRR calculated with one standard deviation.

earmuffs. For an earplug, the data tend to suggest that the premolded devices were not well-fit while the E·A·R earplug was able to achieve a reasonable seal.

For the Bilsom UF-1 earmuff, the mixed-normal model yielded the best fit to the SF REAT distributions at 1000 and 8000 Hz. The distributions did not exhibit the low attenuation mode seen in the premolded earplugs and should not be considered to be bimodal. The mixed-normal model has a range of potential shapes besides the bimodal distribution. Murphy *et al.* (2002) examined the SF REAT data from the NIOSH laboratory and concluded that the mixed-normal model in all cases yielded an accurate fit of the distributions. The mixed-normal model can provide an accurate fit, but not significantly better than the normal or gumbel models.

The data collected for the E·A·R plug demonstrate an increase in the mean REAT as the experimenter involvement increases. The EF data exhibit a notch at 2000 Hz, which may indicate the maximum attenuation of the earplug has been reached (Berger, 1983). The REAT distributions for IUF are more evenly spread across a range of attenuation than the SF distributions below 1000 Hz. This difference may result from the interlaboratory differences observed under the IUF method. One of the laboratories achieved greater REATs for the lower frequencies than other laboratories. The broad range of REATs could reflect the effect of insertion depth and canal size on the ability of the E·A·R plug to achieve a tight seal. Presumably, a deeper insertion of the E·A·R plug yielded the greater REAT for the EF method compared to the IUF or SF methods. In the EF data, none of the REATs at any frequency were less than 7.5 dB.

For the premolded V-51R earplug, the influence of experimenter intervention was clearly evident. The mean values for the EF and IUF REAT histograms were greater than the SF mean REATs at all frequencies. As the stimulus frequency increased above 1000 Hz the differences in mean REATs between the fitting methods decreased. The V-51R earplugs yielded less average attenuation than the foam earplug at all frequencies. Depending upon the quality of fit, V-51R earplugs gave more attenuation than the Bilsom muff for some individuals in the lower frequencies.

Similarly, the premolded EP100 earplug exhibited a low attenuation mode for several frequencies in both the IUF and SF REAT distributions. For 500 Hz, the distribution was not significantly different from normal ($p = 0.052$). Otherwise, the maximum likelihood tests would have identified the distribution as mixed-normal. For the REAT distributions at 4000 and 8000 Hz, a small bulge can be seen in the histogram and is correlated with the larger bulges seen in the lower frequency distributions.

For premolded earplugs, the seal of the ear canal by the plug flanges is the critical factor influencing the bimodal character of the REAT distribution. In Royster *et al.* (1996), the earmuff REAT data were reported for the case where safety glasses were worn by the subjects. The lack of seal continuity around the stems of the glasses reduced the REAT performance. Similarly, for earplugs with an orifice, the insertion loss is about 0 dB for frequencies below 500 Hz (Hamery *et al.*, 1997). As the experimenter involvement increased, the mean REATs increased and more of the distributions were unimodal. A second factor might be the quality of the manufacturer's instructions. The mean IUF REATs were greater when subjects were coached rather than relying only on manufacturer instructions.

For foam earplugs, the attenuation characteristics of the foam and the bone-conduction limits are relatively consistent across subjects (Berger, 1983; Berger *et al.*, 2003). Consequently, varying insertion depths across subjects and the ability to seal the ear canal with a minimum of material probably contribute to the unimodal REAT distributions.

Several root causes of the problem of poor and improper fits could be investigated. The subject may have small ear canals, possess a low tolerance of discomfort, or perceive that deep insertion will damage the tympanic membrane. One may be concerned for the eardrum for a very deep insertion, but the depths typically observed for SF tests were nowhere near that deep. Instructions may be inadequate. The protector's design may not lend itself to a tight seal against the head or ear canal walls. The verification of such speculations would require further analysis of ear canal data and materials used in manufacturing the devices. The clear conclusion is that the type and design of the HPD affect the fit along with the amount of experimenter involvement.

B. Repeatability and reproducibility

The standard deviations for within-subject repeatability and between-subject reproducibility exhibited trends with increasing experimenter involvement and with the spread of data observed in the REAT histograms. For repeatability, the Bilsom muff had the smallest $\sigma_{\text{within-subject}}$ and was followed by the E·A·R plug, the V-51R and EP100 earplugs, respectively. The standard deviations for all devices at 8000 Hz were greater than the value at 4000 Hz. Some slight trend can be seen for increasing standard deviations with decreasing experimenter involvement. The spread of these values was seldom more than 0.2 dB and not more than 0.4 dB for any frequency. The lack of change between fitting methods can be understood as a metric of how subjects fit the device and perform the psychoacoustic task. The smaller standard deviations for within-subject repeatability for the Bilsom device reflect the consistency of fit for repeated tests. For the three earplugs, the occluded threshold measurements may have increased overall variance within subjects. In Franks *et al.* (2003), the variance of repeated occluded thresholds was greater than repeated measures of unoccluded thresholds. Unfortunately, the data examined in this paper were the REAT measures for each subject's trial, not the unoccluded and occluded thresholds.

The standard deviations for the between-subject reproducibility include the additional information about the consistency of fit across subjects. The Bilsom earmuff had the smallest standard deviations which again reflects the ease of fitting the device. The E·A·R plug had smaller deviations than the other plugs. Since fitting the earplugs requires some training, the standard deviations for all the earplugs reflect a trend of decreasing standard deviations with increasing experimenter intervention. The differences in the range are less than 0.5 dB for most devices and frequencies.

The standard deviations for between-laboratory reproducibility yielded results contrary to the within-subject repeatability and between-subject reproducibility. As stated before, the increase in the E·A·R plug IUF between-laboratory standard deviations for frequencies below 1000 Hz has been attributed to greater attenuation achieved by one laboratory. These effects are evident in the increased range of the REATs presented by Royster *et al.* (1996) and are reflected in the values for $\sigma_{\text{between-subject}}$ for the IUF condition in Fig. 5.

One problem with the large between-laboratory standard deviations is the inability to make comparisons of the REATs. Some of the devices and methods failed to produce any between-laboratory standard deviations, $\sigma_{\text{between-laboratory}}$, less than 2 dB. Using three standard deviations as a criterion, differences in REAT for premolded earplugs measured with the SF method only begin to be meaningful when they are greater than 6 dB. The two-lab EF data tend to have standard deviations less than 2 dB and sometimes nearer to one. When the between-laboratory variations are as small as 2 dB, it is appropriate to make comparisons of REATs measured in the different laboratories.

The increased variance at 8000 Hz needs to be addressed since it influences the number of test subjects for the two premolded earplugs and the earmuff. The effect of the acoustic leak at low frequencies correlates with poor attenuation at 8000 Hz. In Hamery *et al.* (1997), the insertion losses of a variety of earplugs were measured for many configurations of the size and position of a nonlinear orifice. Several plots exhibited a maximum insertion loss below 8000 Hz and a local minimum at the 8000-Hz frequency. If the poorly fit protector behaves similarly to the earplug with an orifice, then the correlation of poor attenuation at 8000 Hz might be explained by a resonance of the occluded volume. The protectors tend to have more attenuation at 8000 Hz. These two factors work against having small variances. The protector that can reduce the potential for leakage (earmuffs and foam plugs) will likely have smaller variance and consequently require fewer subjects to achieve a given minimum detectable difference.

C. Power calculations

Lastly, the power calculations need to be examined. The power calculations demonstrate that consistent HPD performance across subjects require fewer subjects to achieve a desired level of resolution (see Fig. 8). For the Bilsom muff, the number of subjects necessary to achieve a 6-dB resolution with the SF method is less than 4. The E·A·R plug requires fewer than 10 subjects. The other two devices, EP100 and V-51R, respectively, require 31 and 22 subjects to achieve the 6-dB resolution. The worst-case frequency is the appropriate choice upon which to estimate the number of subjects. If the confidence level and power were made more stringent (e.g., $1-\alpha=0.9$ and $\beta=0.85$), the minimum detectable difference will increase. While the minimum detectable difference is dependent upon the number of subjects tested, it is also a function of σ_{trial} and σ_{subject} for the device. With power calculation outcomes, it is possible to determine whether a rating for a protector has any relevance when a set number of subjects is used.

The current noise reduction rating provides no estimate other than the standard deviations listed on the secondary label of the error that might be associated with the rating. The definition of the power calculation and development of a method to assess the error associated with the REAT measurement and subsequent NRR rating would achieve parity among ratings. The earmuff data have smaller standard deviations across subjects and laboratories than the earplugs. This difference implies that a worker can expect to achieve attenuations near to what is measured in the laboratory when wearing earmuffs. An error estimate for the NRR of the UF-1 earmuff should be smaller than the error estimate of the earplugs. Such estimates are not currently a part of the EPA rating regulation or the ANSI S12.6-1997 (ANSI, 1997) standard.

D. Subject-fit method and the ANSI S12.6-1997 (ANSI, 1997 standard

This research as described in the Introduction and in Royster *et al.* (1996) was designed by the S12 Working Group 11 to develop a laboratory testing standard that was predictive of real-world protection received by motivated workers. Two protector fitting protocols were investigated. The data derived from the subject-fit method compared to the informed-user-fit method did not yield the expected results. The standard deviations for between-subject reproducibility and between-laboratory reproducibility were expected to increase with decreased experimenter involvement. Instead, the standard deviations for between-laboratory reproducibility fluctuated considerably for the IUF method. Therefore, the more consistent and realistic method with respect to real-world data (Berger *et al.*, 1998) proved to be the subject-fit method.

The authors of the ANSI S3.19-1974 (ANSI, 1974) standard selected a sample size of ten subjects and three test repetitions for each subject. Each repetition consisted of a pairing of occluded and unoccluded thresholds. As demonstrated from the present power calculations for determining the estimated subject sample size for a 6-dB resolution, the working group's selection of ten subjects and two trials was valid for the Bilsom UF-1 earmuff where as few as four subjects with two trials would have been sufficient. As well, ten subjects with two trials would be sufficient for the E·A·R plug for 6-dB resolution. However, testing premolded products requires more subjects and, as the V-51R and EP100 are examples of hearing protectors that are fitted ineffectively, the working group selected 20 subjects with two trials. As Eq. (8) and Fig. 7 show, increasing the panel size increases the statistical power and decreases the minimum detectable difference. As ANSI S12.6-1997 (ANSI, 1997) provides for adequate sample sizes for at least 6-dB resolution for almost any type of hearing protector, it is also possible to calculate the precision of REATs for any set of 10 subjects for earmuffs or 20 subjects for earplugs.

The ANSI S3.19-1974 (ANSI, 1974) statistical treatment of the three repetitions of REAT measurements from ten subjects as 30 statistically independent, uncorrelated data points is incorrect. The correct approach is to average the three repetitions from each subject and then use the individual

averages for calculating means and standard deviations. This statistical shortcoming has been remedied in the S12.6-1997 standard where the averages of repeated measures are used for determining means and standard deviations.

Based upon that rationale, the power calculations for the subject-fit method using 6-dB resolution yield reasonable subject sample sizes for the uniformly performing HPDs (Bilsom muff and E·A·R plug). The power calculations provide a different perspective with which to evaluate the performance of a hearing protector. Those devices which are easily fit and produce highly consistent results across frequencies will yield smaller variances and, consequently, require fewer subjects to achieve a given resolution.

V. CONCLUSIONS

The analysis of the between-subject reproducibility and between-laboratory reproducibility of the devices tested in the four- and two-lab studies has been presented. Some of the underlying causes for increased variance for various protectors have been discussed. The variance of repeatability increases with decreasing experimenter intervention. The variance for reproducibility is most consistent when the experimenter factor is removed. In a separate analysis the Working Group also identified the subject-fit method as the best estimator of field performance (Berger *et al*. 1998), and for that reason the Method-B procedure was initially selected for inclusion in the 1997 standard. The results of the analyses in this report verify that, for purposes of data reproducibility, Method B is an apt choice as well.

ACKNOWLEDGMENTS

The authors wish to acknowledge Dr. Charles Nixon's contributions for his oversight in data collection at Wright Patterson Air Force Base. The authors wish to thank Dr. Doug Ohlin for his assistance with developing the research protocol and support from the U.S Army. Portions of this work were supported by the U.S. EPA Interagency Agreement 75090527.

ANSI (**1999**). S3.6-1999, *American National Standard Specification for Audiometers* (American National Standards Institute, New York).

ANSI (**1974**). S3.19-1974, *American National Standard Method for the Measurement of Real-ear Protection of Hearing Protectors and Physical Attenuation of Earmuffs* (American National Standards Institute, New York).

ANSI (**1997**). S12.6-1997, *American National Standard Method for Measuring Real-ear Attenuation of Hearing Protectors* (American National Standards Institute, New York).

Berger, E. H. (**1983**). ''Laboratory attenuation of earmuffs and earplugs both singly and in combination,'' Am. Ind. Hyg. Assoc. J. **44**, 321–329.

Berger, E. H., Franks, J. R., Behar, A., Casali, J. G., Dixon-Ernst, C., Kieper, R. W., Merry, C. J., Mozo, B. T., Nixon, C. W., Ohlin, D., Royster, J. D., and Royster, L. H. (**1998**). ''Development of a new standard laboratory protocol for estimating the field attenuation of hearing protection devices. Part III. The validity of using subject-fit data,'' J. Acoust. Soc. Am., **103**, 665–672.

Berger, E. H., Kieper, R. W., and Gauger, D. (**2003**). ''Hearing protection: Surpassing the limits to attenuation imposed by the bone-conduction pathways,'' J. Acoust. Soc. Am. **114**, 1955–1967.

Casali, J. G., and Park, M. Y. (**1991**). ''Laboratory versus field attenuation of selected hearing protectors,'' Sound Vib. **10**, 28–38.

Franks, J. R., Murphy, W. J., Johnson, J. L., and Harris, D. A. (**2000**). ''Four earplugs in search of a rating system,'' Ear Hear. **21**, 218–226.

Franks, J. R., Murphy, W. J., Harris, D. A., Johnson, J. L., and Shaw, P. B. (**2003**). ''Alternative field methods for measuring hearing protector performance,'' Am. Ind. Hyg. Assoc. J. **64**(4), 501–509.

Hamery, P., Dancer, A., and Evrard, G. (**1997**). *Étude et réalisation de bouchons d'oreilles perforés non linéaires*, ISL R128/97 (Insitut Franco-Allemand de Recherces de Saint-Louis, Saint–Louis).

ISO 5725-2 (**1994**). *Accuracy (trueness and precision) of measurement methods and results—Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method* (International Organization for Standardization, Geneva).

Murphy, W. J., and Franks, J. R. (**1998**). ''Analysis of repeatability and reproducibility of hearing protector real-ear attenuation at threshold measured with three fitting methods,'' Nat. Hear. Cons. Assoc., 19–21 February, Albuquerque, NM.

Murphy, W. J., and Franks, J. R. (**2001**). ''A reevaluation of the Noise Reduction Rating,'' Meeting of the Commissioned Officers Association of the US Public Health Service, 29 May, Washington, DC.

Murphy, W. J., Franks, J. R., and Krieg, E. F. (**2002**). ''Hearing protector attenuation: Models of attenuation distributions,'' J. Acoust. Soc. Am. **111**, 2109–2116.

Netter, J., Wasserman, W., and Kutner, M. H. (**1990**). *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs*, 3rd ed. (Irwin, Boston), pp. 970–1001.

NIOSH (**1995**). *The NIOSH Compendium of Hearing Protection Devices*, U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Publication No. 95–105.

Rosner, B. (**1990**). *Fundamentals of Biostatistics*, 3rd ed. (Duxbury, Belmont).

Royster, J. D., Berger, E. H., Merry, C. J., Nixon, C. W., Franks, J. R., Behar, A., Casali, J. G., Dixon-Ernst, C., Kieper, R. W., Mozo, B. T., Ohlin, D., and Royster, L. H. (**1996**). ''Development of a new standard laboratory protocol for estimating the field attenuation of hearing protection devices. Part I. Research of Working Group 11, Accredited Standards Committee S12, Noise,'' J. Acoust. Soc. Am. **99**, 1506–1526.

SAS/STAT Software (**1998**). Version 6.12 SAS Institute Inc.

S-Plus 6.1 for Windows (**2002**). Insightful Corporation, Seattle, WA.

U.S. Environmental Protection Agency (**1978**). CFR Title 40, subchapter G, 211, subpart B—Hearing Protective Devices, U.S. EPA.

Separation of concurrent broadband sound sources by human listeners

Virginia Best
Department of Physiology, University of Sydney, NSW 2006, Australia

André van Schaik
Department of Physiology, University of Sydney, NSW 2006, Australia and School of Electrical and Information Engineering, University of Sydney, NSW 2006, Australia

Simon Carlile[a)]
Department of Physiology, University of Sydney, NSW 2006, Australia and Institute for Biomedical Research, University of Sydney, NSW 2006, Australia

(Received 21 August 2002; revised 14 October 2003; accepted 15 October 2003)

The effect of spatial separation on the ability of human listeners to resolve a pair of concurrent broadband sounds was examined. Stimuli were presented in a virtual auditory environment using individualized outer ear filter functions. Subjects were presented with two simultaneous noise bursts that were either spatially coincident or separated (horizontally or vertically), and responded as to whether they perceived one or two source locations. Testing was carried out at five reference locations on the audiovisual horizon (0°, 22.5°, 45°, 67.5°, and 90° azimuth). Results from experiment 1 showed that at more lateral locations, a larger horizontal separation was required for the perception of two sounds. The reverse was true for vertical separation. Furthermore, it was observed that subjects were unable to separate stimulus pairs if they delivered the same interaural differences in time (ITD) and level (ILD). These findings suggested that the auditory system exploited differences in one or both of the binaural cues to resolve the sources, and could not use monaural spectral cues effectively for the task. In experiments 2 and 3, separation of concurrent noise sources was examined upon removal of low-frequency content (and ITDs), onset/offset ITDs, both of these in conjunction, and all ITD information. While onset and offset ITDs did not appear to play a major role, differences in ongoing ITDs were robust cues for separation under these conditions, including those in the envelopes of high-frequency channels. [DOI: 10.1121/1.1632484]

PACS numbers: 43.66.Qp, 43.66.Pn [LRB] Pages: 324–336

I. INTRODUCTION

There are several acoustic cues that define the location of a sound source relative to a human listener [for review see Blauert (1983), Middlebrooks and Green (1991), and Carlile (1996)]. For a sound source located away from the midline, the difference in path length to the two ears results in a difference in the arrival time of the sounds at each ear (interaural time difference, ITD). Furthermore, a sound originating from one side of the head will be more intense in the near ear as compared to the far ear, resulting in an overall interaural level difference (ILD). As the ears are placed more or less symmetrically on the head, the binaural cues are ambiguous; a given interaural difference corresponds to a set of locations on a rough cone centered on the interaural axis. To resolve these ambiguities, the auditory system relies on the three-dimensional asymmetry of the auditory periphery. Location-dependent spectral filtering of sounds by the pinnae, head, and torso has been shown to provide monaural spectral cues to sound source location (Batteau, 1967; Gardner and Gardner, 1973). In addition, the ILD is also frequency dependent and results from differential filtering effects of the head and auditory periphery.

Importantly, sound arriving at the ears rarely comes from a single source. It is of interest to a listener to correctly attribute acoustic information to different sources to aid in the identification and localization of objects in the environment. There exists a large body of research detailing the stimulus factors that assist in auditory grouping and auditory segregation [so-called ''auditory scene analysis:'' see Bregman (1990)]. Auditory grouping occurs on the basis of frequency similarity and spectral continuity (Bregman and Campbell, 1971; Bregman, 1990), and segregation is aided when auditory objects differ in their spectral content or temporal structure (Perrott, 1984; Stellmack, 1994). It has also been known for several decades that the location of competing sound sources has an impact on their separability. The spatial separation of sources improves the detection of tones presented concurrently with noise [see Durlach and Colburn (1978) for review] as well as aiding complex phenomena such as speech intelligibility in noisy environments (e.g., Hirsh, 1950; Dirks and Wilson, 1969). Despite extensive work in this area, only a few studies have attempted to measure systematically the spatial resolution of the auditory sys-

[a)]Author to whom correspondence should be addressed. Electronic mail: simonc@physiol.usyd.edu.

tem for simultaneous stimuli, and to define the role of the different localization cues under these circumstances.

Perrott (1984) defined the concurrent minimum angle (CMAA) as the threshold separation angle required to distinguish two concurrent sounds. He measured the CMAA for sources distributed in a horizontal plane and reported it to be larger than the angle required to determine the direction of displacement of two *sequential* sounds (minimum audible angle, MAA). Perrott presented pairs of tones of different frequency, and asked subjects to judge the relative location of the pair by indicating whether the higher tone was to the left or right of the lower tone. Using a criterion of 75% correct, he measured CMAAs and noted a significant effect of azimuth, with CMAAs of 4°–10° at the front increasing to 30°–45° at a lateral displacement of 67°. Divenyi and Oliver (1989) extended this work to examine more complex sounds including amplitude- and frequency-modulated tones, and found a similar threshold increase for stimuli located at 80° as compared to 0° azimuth. In these studies, the concurrent stimuli differed in their frequency content and so binaural effects could not be isolated from pitch effects. In fact, Perrott (1984) reported that CMAAs increased if the frequency difference between the two sources was reduced, indicating that the two parameters are confounded. Divenyi and Oliver (1989) also noted that, for their more broadband stimuli, spectral overlap was detrimental to resolution.

These studies, and indeed most studies of multiple source perception, have only examined separation in the horizontal plane, where binaural cues are dominant and localization is relatively robust for most stimulus types. In the present study, one objective was to examine sound sources coming from many directions around the listener, displaced both horizontally and vertically. Importantly, broadband noise stimuli were chosen because they are well localized in all dimensions (Carlile *et al.*, 1997) as opposed to narrowband stimuli which are poorly localized in elevation and in the front–back dimension (Butler, 1986; Middlebrooks, 1992). Effectively, in previous experiments, the bandwidth required for accurate localization had been sacrificed for the sake of gaining an adequate frequency separation of two sources. The study reported here aimed to demonstrate how listeners perceive concurrent sounds given *only* spatial separation and *no* differences in source content, i.e., no pitch cues. It is not entirely clear what acoustical cues (for localization and/or separation) are available in the sum of two such stimuli. It is likely that interaural temporal cues related to the two independent sources are present (see Sec. VI C), but the spectrum received at each ear will be a sum of the spectra expected from the pinna filtering of each of the sources.

Most previous studies in this area have required some form of identification or localization of target sounds in the presence of distracting sounds. However, in the present study the competing sounds were indistinct, and hence subjects were asked to simply indicate whether they perceived sound arriving from a single location or from two distinct locations [this is how we define ''separation;'' it has also been called ''detection of spatial separateness'' (Noble *et al.*, 1997)]. However, our ''simple'' task is quite difficult for two reasons. First, the two sources presented are identical in their long-term spectrum and temporal characteristics and so cannot be segregated on the basis of their content or identity. Second, as these sounds are broadband, they both produce global activation along the basilar membrane and thus a complete sharing of receptors must take place. Separation cannot commence at the periphery as is the case with differing tones, which are processed in distinct frequency channels, but must rely on more central processing of spatial cues.

The first experiment aimed to examine the ability of listeners to resolve the simultaneous pair and compared this ability in different spatial regions around the listener. The following experiments examined in some detail the contribution of different localization cues to this perception, with a focus on the possible role of ITD. The ITD cue has two main components: a transient onset (and offset) ITD and an ongoing ITD. For pure tones the ongoing ITD refers to cycle-by-cycle phase differences, and is only useful at low frequencies (below about 1.5 kHz) where periodicity information is preserved by phase-locking neurons in the auditory pathway (see Moore, 1997). For complex waveforms, however, it is known that interaural delays in the slowly varying envelope of high-frequency channels (above 1.5 kHz) can also be processed by the auditory system (Henning, 1974, 1980; McFadden and Pasanen, 1976). Lateralization studies have shown that ongoing ITD is a more potent indicator of lateral position than is onset/offset ITD, with the latter only having a strong influence when the signal is brief [shorter than about 10 ms (Tobias and Schubert, 1959)] or the ongoing ITD is ambiguous (Kunov and Abel, 1981; Buell *et al.*, 1991). In the following experiments, separation of concurrent noise sources was examined upon removal of low-frequency content (experiment 2a), onset/offset ITDs (experiment 2b), both of these in conjunction (experiment 3a), and all ITD information (experiment 3b). These manipulations were made possible by exploiting a virtual auditory space stimulation technique.

II. METHODS

A. Subjects and environment

Four subjects (two males and two females, aged 23–33 years) participated in the experiments. All had normal hearing by audiometric testing and some experience in auditory psychophysical testing. Experiments consisted of a series of short tests, during which subjects were seated in a darkened, sound-attenuating chamber. Stimuli were presented in virtual auditory space (see Sec. II B) over in-ear headphones at approximately 50 dB sensation level. A small amount of training was conducted prior to testing, and this is described in Sec. II D.

B. Stimulus generation

Virtual auditory space (VAS) refers to a spatial auditory percept that is created using headphone presentation. For the presentation of multiple concurrent sounds at finely spaced locations this was an essential tool, as it avoided the need for multiple speakers and an extremely sensitive placement system. In addition, some of the stimulus manipulations used

would simply not be achievable using stimuli presented from loudspeakers in ''real'' space. Individualized VAS is generated by accurately simulating the wave pattern at the eardrum occurring after free-field stimulation with an external sound source (Carlile, 1996). For different sound source locations in space relative to a listener's head, a different pattern of filtering is imposed on the sound by the head and outer ears (head-related transfer function, HRTF). These functions are recorded routinely in our laboratory by the placement of microphones in the ear canal to record the impulse response to a specific broadband stimulus [blocked-ear canal recording (Middlebrooks *et al.*, 1989; Moller *et al.*, 1995]. The recording stimulus was a 1024-bit Golay code pair presented 12 times, with the resultant input averaged to increase the signal-to-noise ratio (Zhou *et al.*, 1992). Subjects were seated in an anechoic chamber, within which a hoop system carrying a loudspeaker could be moved to position the loudspeaker at any location on an imaginary sphere (radius 1 m). The subject's head was positioned in the center of this sphere by fixing a laser from the front (0°,0°) on the nose, and a laser from the left side (−90°,0°) on the ear canal entrance. Subjects were instructed to keep their heads in this position for the duration of the recording process (about 30 min) with the assistance of a chin-rest. To minimize artifacts that may arise as a result of small head movements during the Golay code pair repetitions (Zahorik, 2000), head position was continuously monitored using an electromagnetic head-tracker (Polhemus) and recording stopped immediately when deviations occurred. Subjects were instructed on how to readjust their heads using a small LED display that signaled the direction of deviation relative to the calibrated position. Impulse responses were obtained for 393 positions evenly spaced on the sphere. This procedure has been described in more detail elsewhere (Pralong and Carlile, 1996).

Microphone (Sennheiser KE 4-211-2) and system transfer functions were removed from the impulse response functions by deconvolving the microphone transfer function, and HRTFs were extracted. Location-independent components were removed to leave only the directional transfer functions (DTFs) for each ear (see Middlebrooks *et al.*, 1989). The two DTF filters corresponding to a particular location could then be convolved with any sound stimulus and delivered via in-ear headphones (Etymotic Research ER-2) to the subject to give rise to a virtual externalized stimulus at that location. Analog-to-digital and digital-to-analog conversion for recording and playback occurred at a sampling rate of 80 kHz [Tucker-Davis Technologies (TDT) System II] and stimuli were delivered to an amplifier via a programmable attenuator (TDT: PA4). Stimulus synthesis and delivery as well as data recording and visualization made use of MATLAB 5.3 software (Mathworks Inc.).

Two tests were undertaken to ensure that listeners were valid subjects for VAS experimentation. First, subjects were required to confirm subjectively that their individualized VAS stimuli were realistic and externalized. Second, a standard localization test was performed to ensure that listeners could localize with a suitable level of accuracy in both the free-field and in VAS [see Carlile *et al.* (1997) for details]. Briefly, this test required a subject to stand in darkness in the center of the anechoic chamber and indicate his/her perceived location of a series of 150-ms broadband noise bursts (300 Hz to 16 kHz). These were presented randomly from a subset of 76 recording positions, and responses were indicated using a nose-pointing technique and recorded via the head-tracker. All subjects used in this study were found to be accurate localizers in both the free-field and in VAS, and a summary of their localization data can be found in the Appendix.

Importantly in this study, it was required that stimuli be presented at often closely spaced locations on the virtual sphere of space. However, in the recording process, only a discrete set of 393 locations were obtained distributed on the sphere at approximately 10°–13° intervals. Thus a spatial interpolation procedure, based on spherical thin-plate splines and principal components analysis, was employed to allow locations *within* this grid to be simulated (Carlile and Leung, 2001). It has been confirmed in the laboratory that this procedure produces DTFs that are statistically identical to equivalent measured DTFs (in terms of the structure of the filters), and that virtual sound stimuli based on these DTFs are localized as well as those based on measured ones (Carlile *et al.*, 2000).

To generate stimuli imitating concurrently activated sound source pairs, two independent random noises were filtered with DTFs corresponding to the two desired locations and then added together. For ''zero separation'' stimuli, the same procedure was followed but the two DTF pairs used were identical.

C. Psychophysical procedures

The experiment took the form of a single-interval forced choice procedure (see Sec. II D for a short discussion of this approach). Subjects were presented with a stimulus pair and were asked to indicate, by pressing one of two buttons, whether they perceived sound to be coming from one or two source locations.

Stimulus locations are described throughout the study using azimuth and elevation with respect to a single-pole coordinate system. For experiments 1, 2a, and 2b (see below) testing took place at five reference locations on the 0° elevation horizontal plane: 0°, 22.5°, 45°, 67.5°, and 90° azimuth (Fig. 1). In each trial, one stimulus in the concurrent pair was presented from a reference location and the other from a test location displaced in either azimuth (horizontal separation) or elevation (vertical separation) or from the same location. Note that binaural cues change with horizontal separation (maximum rate-of-change at the front, minimum rate-of-change at the side) and also with vertical separation using the single-pole coordinate system (maximum rate-of-change at the side and *no* change at the front). For each reference location, 15 test locations were chosen on the basis of preliminary testing. For vertical separation, the testing range was the same for each of the five locations, and spanned the entire available range of elevations: from −45° to 90° (directly overhead). Ranges for horizontal separation varied with location as required to cover a suitable range: ±21° for 0° azimuth; ±32° for 22.5° azimuth; ±42° for 45° azimuth; ±53° for 67.5° azimuth; and ±63° for 90° azimuth.

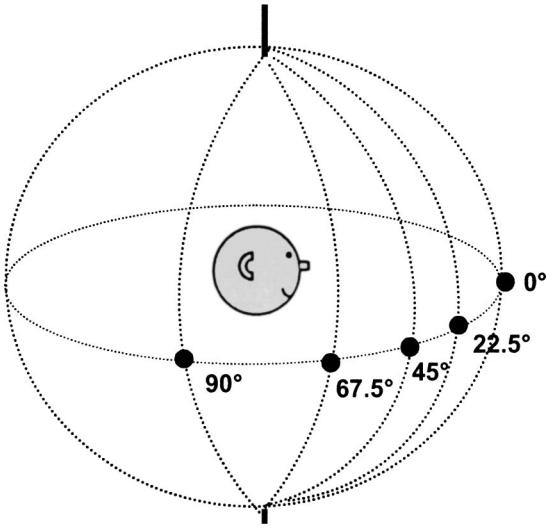


FIG. 1. The single pole coordinate sphere used to define locations in space. The black dots illustrate the five reference locations examined: 0°, 22.5°, 45°, 67.5°, and 90° azimuth, all on the 0° elevation plane at the level of the ears.

All reference location/separation combinations were presented ten times each in a random order, giving 1500 trials in an experimental block (5 reference locations, 2 directions, 15 test separations, 10 repetitions). These were broken down into ten tests of 150 trials, each lasting about 15 min. For experiments 3a and 3b testing only took place at the most frontal reference location (0° azimuth). The same test separations were used for this location as in the other experiments (vertical: 15 values between −45° and 90° elevation; horizontal: 15 values between −21° and 21° azimuth). These shorter experiments each comprised 300 trials (two tests of 150 trials).

Results from each block of experiments were analyzed in turn. Results from a particular reference location were pooled and sorted and psychophysical curves for horizontal and vertical separation were plotted. These illustrate, for each test separation, the percentage of times (out of ten repetitions) that the subject responded that he/she perceived two sources.

D. Response criterion, training, and controls

In establishing the experimental protocol for examining the spatial perception of concurrent stimuli, the widely used two-interval discrimination task was trialed; it required subjects to choose which interval contained two sources and which contained only one. However, this approach was quickly rejected because it was clear that comparisons could be made on the basis of a number of available cues (such as timbral changes), and not necessarily the ones of interest in this study (spatial cues). The single-interval paradigm was thus adopted in order to best measure the effects of spatial separation on the perception of concurrent sources. When presented with a particular stimulus, subjects were required to respond as to whether they perceived one or two source locations. Although this approach did not *ensure* that responses were related to a clear percept of two sources, it did encourage listeners to use cues that were (subjectively) spatial in nature.

One difficulty associated with the single-interval subjective task is that response biases (such as pressing one button more often than the other) cannot be effectively removed as they can in a discrimination task. Only if this bias is consistent within a subject can comparisons across conditions be made with confidence. Another potential problem with the one-interval task is its subjectivity: the experimenter must rely on subjects to adopt a criterion for responding and to maintain this criterion throughout testing. These potential difficulties were dealt with in two ways. First, subjects underwent a small amount of training to stabilize their performance before commencement of data collection. Each subject was run through two tests of the format described in Sec. II C, allowing him/her to become familiar with the stimulus and task and to establish a comfortable criterion. Second, throughout an experimental block, a specific repeated control was included in each test to ensure that a particular subject's bias and criteria did not fluctuate. This control consisted of a small set of 30 trials, which were interleaved with the 150 experimental trials. These were five repeated trials at six horizontal separation values from the 0° reference location set (separations of 0°, 3°, 6°, 9°, 12°, 15°). Responses to these trials provided a measure of response bias and criterion for each test. At the end of an experimental block, the control sets were examined to confirm that they were stable across testing. If a set deviated greatly from the rest, as determined by the experimenters on the basis of informal monitoring, the test from which it came was repeated. Only two tests had to be repeated on this basis (in one subject).

Despite the good within-subject consistency, the fact that response bias and criteria varied across individuals meant that it was difficult to pool responses across the population. Thus for this study the emphasis is not on the absolute value of responses, but rather the pattern of performance of each subject across the various stimuli and test locations.

III. EXPERIMENT 1: BROADBAND STIMULI

A. Stimuli

Experiment 1 was conducted to examine the ability of subjects to separate concurrent broadband sounds. Random noises containing frequencies from 300 Hz to 16 kHz were used (as for localization testing, see Sec. II B). For a given stimulus, two independent noises were filtered with appropriate DTFs and ramped by applying a raised cosine to the first and last 10 ms. The two binaural signals were then added to create concurrent stimuli as depicted in Fig. 2(a).

B. Results

Psychophysical curves for the first experiment are shown in Fig. 3. Data are displayed for individual subjects [Figs. 3(a)–(d)] and mean data are shown in Fig. 3(e). The left-hand panels show data for concurrent sources that were separated in azimuth (horizontally) about the reference location and the five subplots in a panel show data for each of the

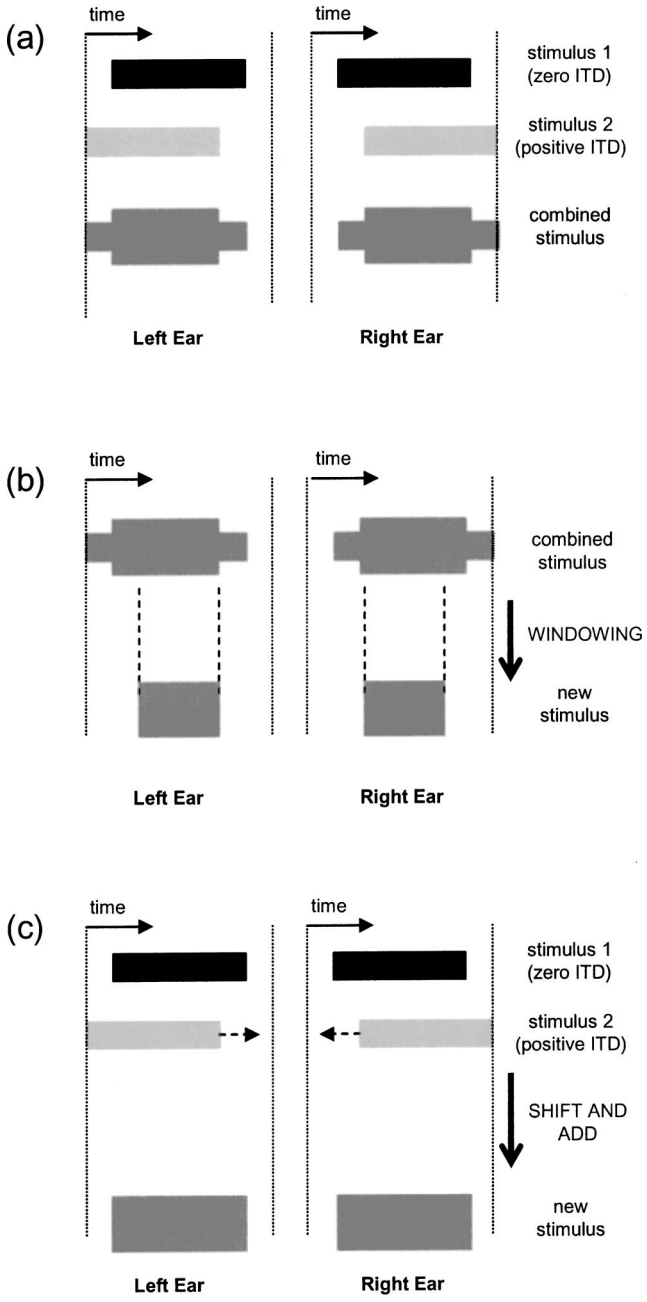


FIG. 2. Temporal features of stimulus manipulations used in experiments 1, 2b, and 3b. The examples given are for a reference stimulus located frontally and a test stimulus at a single lateral location. Left and right ear stimuli are shown in left and right columns. (a) Experiment 1: To create a concurrent pair, the reference stimulus (zero ITD—black bars) is added to a laterally displaced sound (positive ITD—light gray bars) to create a new stimulus. The ITD difference creates two onsets and offsets in each ear, as well as a phase difference in the ongoing portion of the signals. (b) Experiment 2b: The combined stimulus just described is windowed identically in each ear to remove the onset and offset cues. The ongoing phase differences remain in the signal. (c) Experiment 3b: To remove all ITD differences between the two stimuli, the laterally displaced stimulus (light gray bars) is shifted in time in each ear before adding to align it to the zero ITD stimulus. This removes both onset/offset and ongoing ITD differences, but individual spectral characteristics are preserved.

five reference locations. The abscissa shows the separation of the test stimulus with respect to the reference, where negative values indicate leftward separation and positive values indicate rightward separation. Plotted on the ordinate is the percentage of trials (out of 10) where the subject reported that he/she perceived two sources in the presentation.

A feature that is evident for all subjects (and the mean data) is that the psychophysical curves for horizontal separation broaden as the reference position is displaced more laterally, indicating a decrease in spatial sensitivity with increasing laterality of the sources (left panels, Fig. 3). For the most frontal reference location (0°,0°), 100% response rate was reached with much smaller separation angles than for the most lateral location (90°,0°). All subjects showed a similar pattern of results in this experiment, although there were marked differences in overall response levels. For example, subject 4 responded readily to the detection of both sources, as indicated by the relatively narrow troughs [Fig. 3(d)], while subject 1 tended to give more conservative responses [Fig. 3(a)]. This individuality is examined in Sec. VI A. A feature of the data for all subjects is an asymmetry in the psychophysical curves for 67.5° reference location with a drop in response rate as test stimuli were moved towards the back (positive separation). An explanation for this is given in Sec. III C below.

The right-hand panels of Fig. 3 show data plotted in a similar way, but for sources separated vertically about the reference location. Again each of the five reference locations are plotted in separate subplots, and the axes are as described above, except that the separation was in elevation (vertical), with negative values indicating downward separation and positive values indicating upward separation. In these data, the opposite trend can be seen, in that the curves appear to get slightly sharper at the more lateral locations. By far the most striking feature of these data is the frontal midline location. Subjects 1 and 3 show essentially flat psychometric functions indicating that they never (subject 1) or rarely (subject 3) perceived both sources in the pair. Subject 2 gave more positive responses to the stimulus pairs overall but the curve is still relatively flat, indicating that responses were not related to separation. Responses from subject 4 were inconsistent, but he seemed to be able to resolve the sources at large positive separation values. An examination of this subject's HRTFs revealed a marked asymmetry for vertical midline positions, resulting in an elevation-dependent ITD. Specifically, the difference in ITD between elevations of 0° and 90° was measured to be 0.08 ms (compared to the mean value of 0.01 ms for the other three subjects) and this difference is indeed large enough to be a useful cue for separation (see Sec. VI B and Fig. 6).

C. Discussion

The trends seen in the data as the spatial location of testing was varied point strongly to a role for binaural cues in this particular task. The data obtained for horizontal separation on the 0° elevation plane are certainly consistent with this idea. As a result of the position of the two ears on this plane, binaural cues change approximately as a sine function of azimuth, with the rate-of-change being maximal at 0° azimuth and decreasing towards 90° azimuth. As a result, performance in spatial discrimination tasks (Mills, 1972) and absolute localization (Carlile *et al.*, 1997) become poorer in

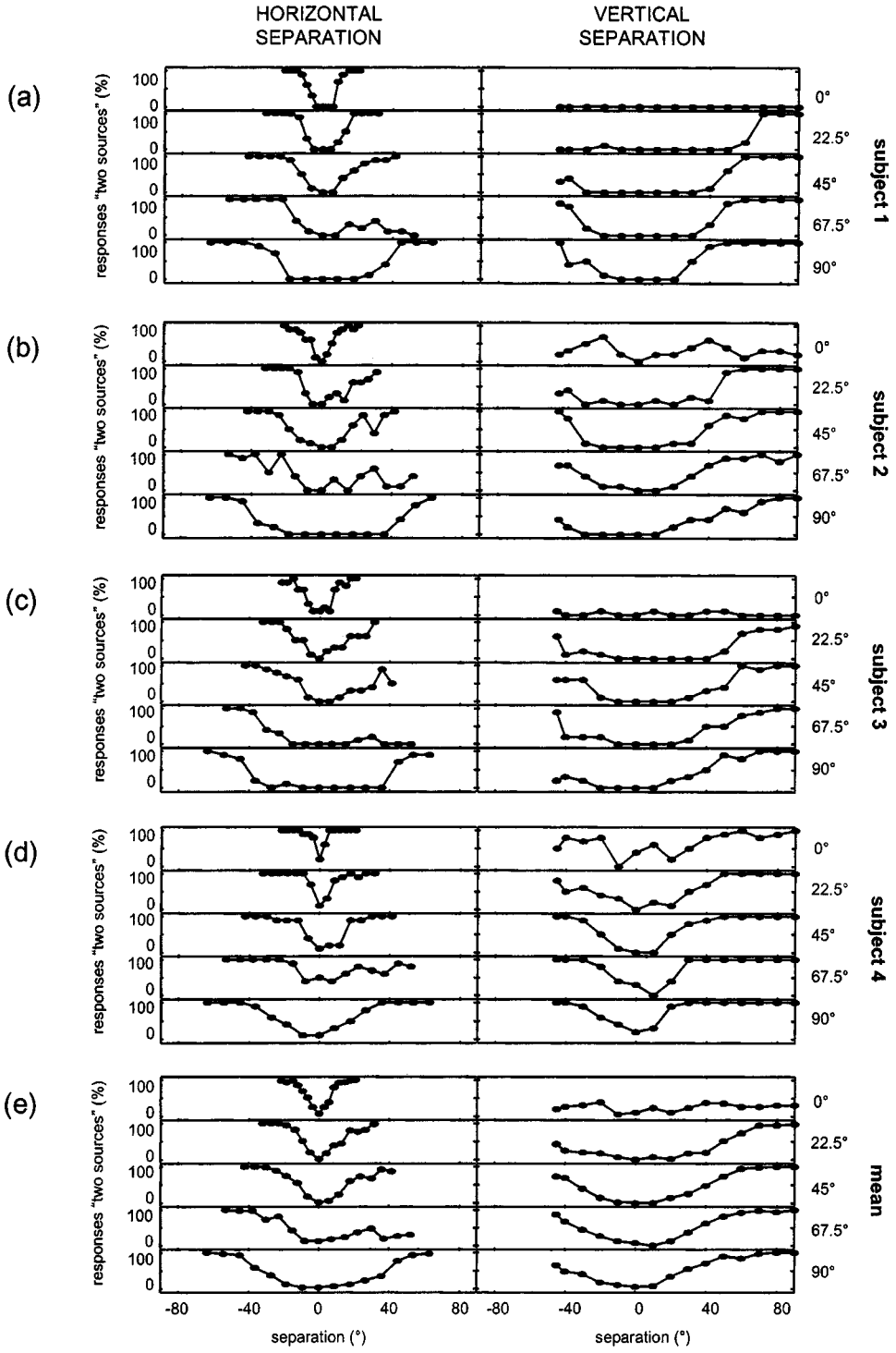


FIG. 3. Psychophysical curves for experiment 1. (a)–(d) show data for subjects 1–4, respectively, and (e) is mean data. The left-hand column shows horizontal separation data and the right-hand column shows vertical separation data. The five subplots in each panel show data for each of the five reference locations (top to bottom: 0°, 22.5°, 45°, 67.5°, and 90° azimuth). For the given test separations (abscissa), the curves show the percentage of trials in which the subject perceived two sources (ordinate).

the horizontal dimension as azimuth increases. In addition, it follows that if separation is dependent on differences in the binaural cues between a pair of stimuli, then the *angle* required to achieve adequate binaural separation would increase with laterality. This is seen in the present data, where the angle of separation required to reach a given response rate increased at the more lateral positions. Consistent with this, Perrott (1984) reported that the concurrent minimum audible angle of tones increased at lateral positions, and Divenyi and Oliver (1989) found a similar effect for amplitude- and frequency-modulated tones.

A special case can be seen at the 67.5° reference position when the test stimuli were separated towards the back. The value of the ITD and ILD for the test stimulus increases as the test location is moved towards 90° azimuth (separations of 9° and 18°) but then *decreases* back towards the reference value when the test location passes 90° (separations of 27° and 36°). When the test location is 112.5° azimuth (separation of 45°), this is the reflection of 67.5° azimuth about the interaural axis and so the binaural cues for the test and reference location are near to identical. Thus these data support the notion that the task is dependent on binaural differences between the two sources.

The pattern of responses seen with vertical separation

also suggests that the rate-of-change of the binaural cues is, at least in part, responsible for performance. At the most lateral reference location (90° azimuth), vertical separation of a concurrent stimulus causes a greater change in binaural cues than at more frontal locations, as the displacement is towards the far ear and away from the near ear. Furthermore, at this location, a given separation in the vertical direction changes the binaural cues by approximately the same amount as the same separation in the horizontal direction. Consistent with this, the psychophysical curves at 90° azimuth are similar for horizontal and vertical separation.

Perhaps the clearest demonstration that binaural cues are important for this task is the fact that when stimuli were separated along the vertical midline, subjects generally perceived only one source (or were confused). In other words, given only one set of binaural cues, the auditory system assumes one source. The implication here is that spectral cues are not sufficient under these circumstances to indicate the presence of the distinct sources. This may, at first, seem surprising because spectral cues are responsible for an accurate ability to localize single sources on the vertical midline: average errors are approximately 5° at (0°,0°) for broadband noise (Carlile *et al.*, 1997; Butler *et al.*, 1990). In addition, Perrott and Saberi (1990) reported that subjects could detect a change in location between sequentially presented click trains when they were separated by 3.5° along the vertical midline. Clearly, however, when a pair of broadband sounds is presented concurrently, as in the current experiment, the individual spectral cues become less useful as they sum at each ear. This may not be the case for more band-limited sounds, where there is less spectral overlap.

As discussed in the Introduction, it is known that ITDs are useful for separating concurrent sources under many circumstances. When two sounds are presented concurrently from different locations, there is a different arrival time at each ear for each source. This results in two different ITDs, and it is feasible that these may be recovered by the auditory system and used to separate the two sounds (see Sec. VI C). It is less clear, on the other hand, how useful relative ILD could be for indicating the presence of distinct sources, as signal energy would be added at each ear, producing a *single* new ILD in any frequency channel. As ITD was the most likely candidate for the trends seen in the data from experiment 1, experiments 2 and 3 were undertaken to confirm and characterize its role.

IV. EXPERIMENT 2: EFFECT OF REMOVING SINGLE COMPONENTS OF THE ITD

A. Stimuli

Experiment 2 was conducted in the same way as experiment 1, but with two stimulus manipulations to remove a component of the ITD cue. In experiment 2a, stimuli were high-pass filtered in order to remove low-frequency ongoing timing information. Filtering was performed in the Fourier domain by zeroing the magnitude of all components below 2 kHz and taking the inverse transform to return to a time domain waveform. In experiment 2b, the aim was to remove onset and offset ITD information, and this was achieved by adding the two DTF-filtered noises together and then windowing the resultant left and right ear stimuli such that they ramped on and off at exactly the same points in time. Specifically, 2 ms was trimmed from each end of the stimuli, which was enough to ensure the removal of the two onset and offset cues that were previously available [Fig. 2(b)]. Because 4 ms of signal was lost in the windowing process, the original sounds were generated with duration 154 ms to give rise to a final 150-ms signal. Technically, this manipulation sets the onset and offset ITDs to zero, as opposed to removing them. Nonetheless, it provides the opportunity to observe the effects of removing the location-dependent variation in this cue.

B. Results

Figure 4 shows data for high-passed stimuli (solid lines) and stimuli with onsets and offsets removed (dashed lines) as compared to the broadband data described in experiment 1 (dotted lines). The same set of separations was tested in this experiment as in experiment 1; however, the symbols have been omitted from this figure for clarity. The curves represent the percentage of responses in which the subjects responded that they heard both sources in the concurrent pair. Results for subjects 1–4 are shown [Figs. 4(a)–(d)] as well as the mean [Fig. 4(e)].

It can be seen for each individual that there was no substantial effect of removing the onset and offset ITD cues on performance; the curves (dashed lines) are generally in agreement with the broadband condition (dotted lines). This is true for both horizontal (left panels) and vertical (right panels) separation and is confirmed in the mean data.

It can also be seen that for subject 4, behavior did not change substantially as a result of the high-pass filtering. The curves (solid lines) for this subject generally agree with the broadband condition (dotted lines), with minor deviations. For subject 2 this is generally true also, although some larger deviations can be seen particularly for horizontal separation at the more lateral locations. For subjects 1 and 3, the high-pass filtering had a more obvious impact on performance, but again only at the more lateral reference locations. Here the psychophysical curves became flatter, and 100% separation rate was reached far less often than in the broadband condition. In an examination of the mean data [Fig. 4(e)] it appears that the front-most reference location (0° azimuth) was not affected by the high-pass filtering. However, some flattening of the curves is apparent at more lateral locations (especially 67.5° and 90° azimuth), representing a decrease in sensitivity to the task.

C. Discussion

It was interesting to find that the separation of concurrent sounds at the frontal location was not impaired by high-pass filtering the signals. It seems that in this situation, the ongoing low-frequency ITDs are not useful, or are redundant, and the perception is maintained by high-frequency acoustical cues in this case. It has been shown previously that high-pass filtering does not significantly affect the accuracy of localization of *single-source* stimuli (Butler and Hu-

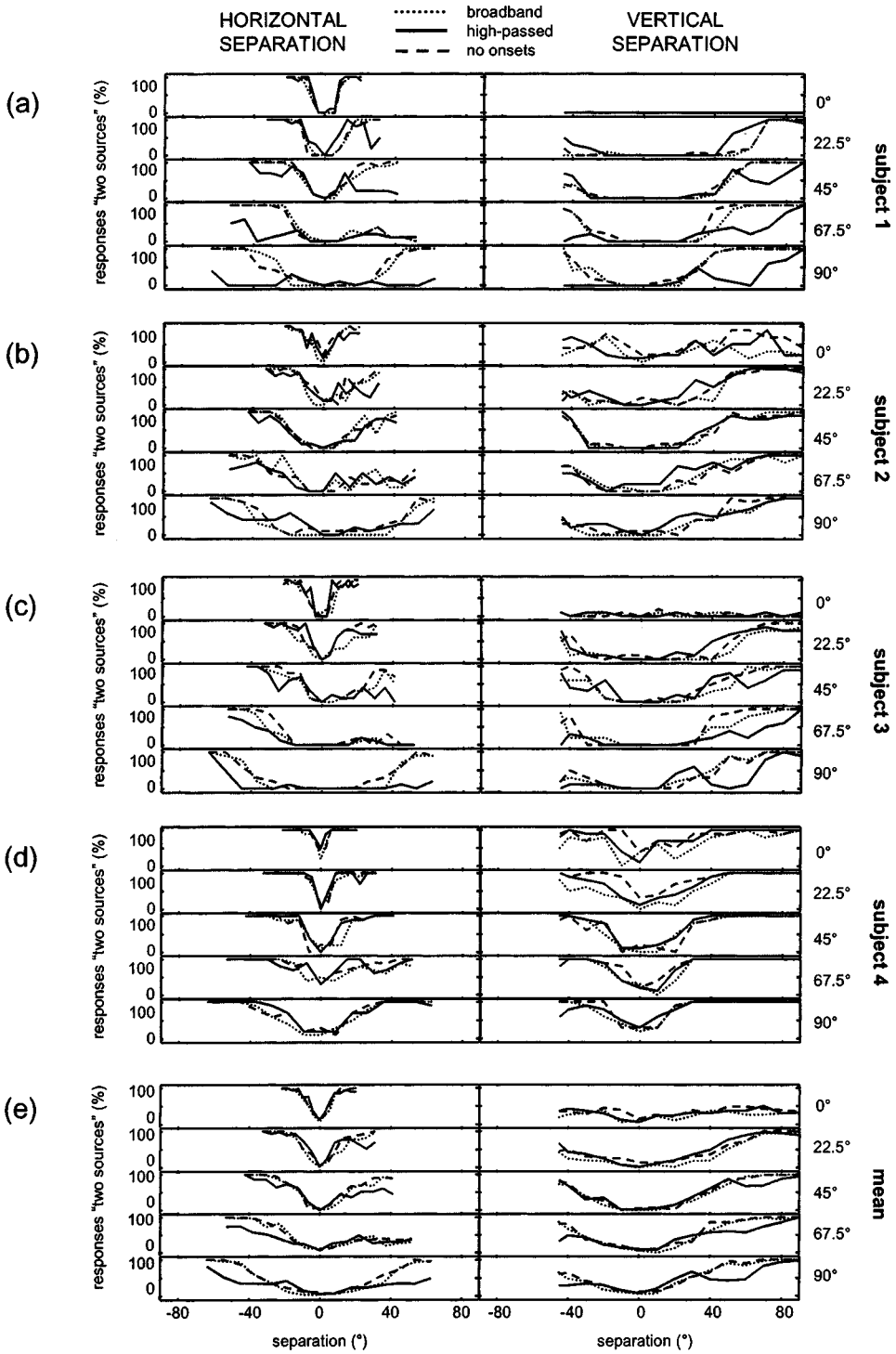


FIG. 4. Psychophysical curves for experiment 2. (a)–(d) show data for subjects 1–4, respectively, and (e) shows mean data. Results are shown for broadband stimuli (dotted lines; same as Fig. 3), stimuli high-pass filtered at 2 kHz (solid lines), and stimuli with onset/offset cues removed (dashed lines). All other details as for Fig. 3.

manski, 1992; Carlile and Delaney, 1999), and the present findings suggest that high-frequency cues can also be sufficient for identifying that more than one location is present in a *multiple-source* stimulus. In some subjects, high-pass filtering was seen to disrupt performance at the more lateral locations. Indeed the task was performed more poorly at these locations in the broadband condition, but with high-passed stimuli performance in some subjects dropped severely. In fact, for one subject the response rate dropped to around zero for *any* stimulus pair presented at the most lateral location [subject 1, Fig. 4(a)]. It is possible that this is related to ITD discrimination thresholds in this region, and this is discussed further in Sec. VI B.

Another finding from experiment 2 was that the presence of ITDs at the onset and offset of the stimulus was not necessary to match the performance of subjects seen in experiment 1. This is somewhat surprising considering that many neurons in the auditory pathway respond preferentially to the onset of a sound stimulus (Pickles, 1988). It is also curious because the onset cue is available and consistent across *all* frequency channels. Indeed for very short impulsive sounds such as clicks, it is the only cue available, and is crucial for lateralization (Tobias and Schubert, 1959). For longer duration stimuli under anechoic conditions, the ongoing phase cue is likely to be a more reliable cue because it can be integrated over time to improve the estimate. It seems

that the transient offset and onset disparities are redundant in the presence of this very robust cue. A similar redundancy of onset ITD information has been shown for the lateralization of ongoing noise bursts presented over headphones (Tobias and Schubert, 1959).

V. EXPERIMENT 3: EFFECT OF REMOVING ITD COMPONENTS IN COMBINATION

A. Stimuli

Experiment 3a was a combination of the two conditions described in experiment 2, where stimuli were high-pass filtered at 2 kHz *and* left and right ear signals were each windowed as just described. In this way, onset and offset ITDs as well as ongoing low-frequency ITD information was removed. This allowed an examination of the usefulness of *high-frequency* information for this separation task, preserving only high-frequency ongoing ITD as well as level and spectral cues.

In the final experiment (experiment 3b) the aim was to remove *all* timing information. To achieve this, the two DTF-filtered stimuli were first created as for condition 1. Following this, the left and right ear signals representing the test stimulus were shifted in time (if required) such that their onsets matched those of the left and right ear signals of a stimulus corresponding to location (0°,0°). The signals were then ramped and added [see Fig. 2(c)]. The result of this manipulation was that the onsets and offsets of the stimulus pair coincided, and the ongoing time differences were matched too. Thus the timing information in isolation corresponded to only one location in space (0°,0°), but the spectral and interaural level information from each individual source was still contained in the composite signal.

These conditions were tested only at the frontal location (0°,0°), as results from the previous conditions had indicated that this location was the only one where no degradation was seen using the ITD manipulations individually.

B. Results

Each panel in Fig. 5 contains data from the previous experiments, as well as data from experiment 3, for horizontal separation at the frontal location only. Individual data for subjects 1–4 are shown [Figs. 5(a)–(d)] as well as the mean data [Fig. 5(e)]. Experiment 1 (dotted lines), experiment 2a (thin solid lines), and experiment 2b (thin dashed lines) have already been discussed, and again the relatively narrow troughs can be seen, indicating a good level of performance in this testing range. The thicker lines show data from experiment 3: high-pass filtered stimuli with onset/offset cues removed (thick solid lines) and stimuli with all ITD cues removed (thick dashed lines).

It can be seen that the curves from experiment 3a (thick solid lines) are very similar to the previous data, indicating that the combination of stimulus manipulations from experiment 2 did not degrade performance at this location. In fact, it seems that in this condition subjects tended to report *more often* that they perceived two sources.

Psychophysical curves for experiment 3b (thick dashed lines) show large individual differences, but it can clearly be seen that the consistent performance obtained in all other conditions was disrupted in this condition. Across all spatial configurations, the number of trials where both stimuli in a concurrent pair were reported was lower than in the other conditions. This is true for the individual subjects [Figs.

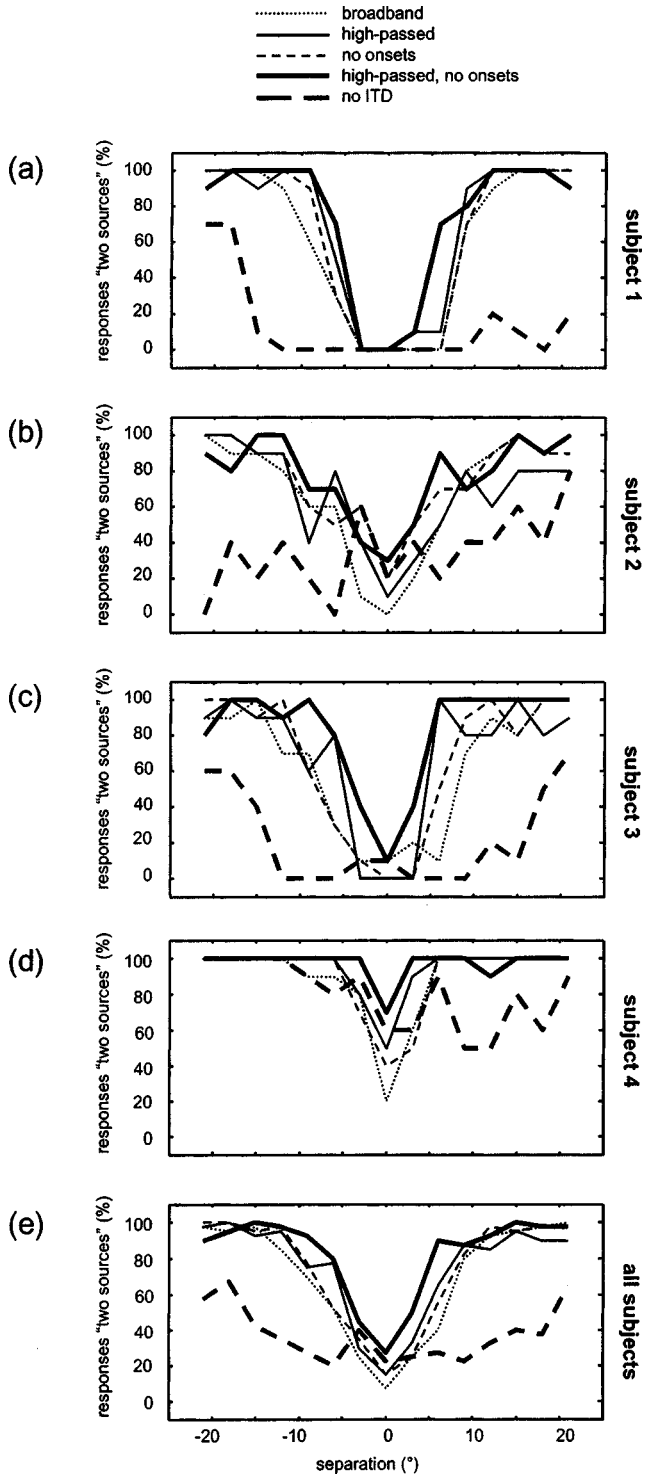


FIG. 5. Psychophysical curves for horizontal separation at the frontal location only, for all experimental conditions. (a)–(d) show data for subjects 1–4, respectively, and (e) shows mean data. Data shown again for experiments 1, 2a and 2b (thin dotted lines, thin solid lines and thin dashed lines) as well as experiment 3a (high-pass filtered and onset/offset cues removed; thick solid lines) and experiment 3b (all ITD information removed; thick dashed lines). For the given test separations (abscissa), the curves show the percentage of trials in which the subject perceived two sources (ordinate).

5(a)–(d)] and for the mean [Fig. 5(e)]. Even for the largest separations, response rate never reached 100% for subjects 1–3. Subject 4 showed a marked performance asymmetry, displaying a reasonable ability to distinguish the two sources with leftward separation, but a disrupted performance with rightward separation.

C. Discussion

Results from experiment 1 suggested that binaural differences drove the separation of concurrent broadband noise sources in the four subjects examined. In experiment 2a, in which the stimuli were high-pass filtered, it was found that low-frequency phase information was not crucial to perform this task, although it aided performance at the more lateral locations. In experiment 2b, in which onset and offset cues were eliminated, it was seen that onset and offset ITD information was not critical and did not even appear to contribute to the performance seen in experiment 1. As there was a possibility of redundancy between these two sets of cues (low-frequency ITDs and onset/offset ITDs), they were both eliminated in experiment 3a to assess their combined importance. At the frontal location, performance was not disrupted under these conditions. Remarkably, subjects could determine the number of sources in a frontally presented signal with this highly impoverished set of localization cues. In the final experiment (experiment 3b), all ITD information was effectively removed by aligning the left and right ear signals in time for each member of the concurrent stimulus pair. The dramatic change in performance in this condition demonstrated that despite the redundancy in the cues, *some* ITD information is crucial for this task. It seems that interaural level differences (ILDs) are ruled out as major cues for this task, as they were available but did not maintain performance in this experiment. This suggests that in experiment 3a, subjects were reliant on ITDs in the envelope of the ongoing high-frequency signal. This is consistent with previous findings that envelope ITDs in high-frequency channels can be useful for sound lateralization (Henning, 1974, 1980; McFadden and Pasanen, 1976). The surprising point is that this cue *alone* produced performance levels equal to those seen when robust low-frequency ITD information was available. This is unexpected, especially because several studies have demonstrated the high-frequency ITD to be a much weaker cue than the low-frequency ITD for lateralization (Yost, 1976; Bernstein and Trahiotis, 1982).

An examination of Fig. 5 reveals that for experiment 3b, there were some effects at larger separations, with subjects reporting the perception of two sources despite the absence of ITD cues. It is possible that this is an artifact of the conflicting ITD and ILD cues at these extremes, where the ITD was zero but the overall ILD was nonzero as one of the source locations was displaced from the midline. Across subjects the mean overall ILDs at the extremes were measured to be 4 dB (test azimuth $-21°$) and -4.5 dB (test azimuth 21°). Indeed it was reported by subjects that some stimuli in this experiment produced an ''unnatural'' percept, and this phenomenon has been reported previously for stimuli where interaural parameters are in conflict (e.g., Gaik, 1993).

VI. GENERAL DISCUSSION

A. Subject performance

In an overview of this collection of data, it is apparent that there are strong individual differences between subjects. As discussed in Sec. II D, efforts were made to ensure that each subject behaved consistently across trials, but as expected the response criterion adopted by a particular subject was quite individualized. This individuality showed up mainly in the overall tendency of subjects to respond that they perceived two source locations in a stimulus presentation. In order to quantify this and inspect the individual differences, a ''false alarm'' rate was calculated from the trials in which both noise sources emanated from the same location (''zero separation'' trials). Across the entire testing set there were 250 of these trials, and for each subject the total percentage of false alarms was calculated. For subjects 1–3 this rate was very low (0%, 6%, and 1.5%, respectively) but for subject 4 the occurrence was higher (26%). This indicates that subject 4 had a greater tendency to respond positively to two source locations, i.e., a less strict response criterion. However, despite these intersubject differences, the overall patterns in the data were the same across subjects.

Furthermore, it is clear from the data that subjects could resolve the pair of noises with a high degree of certainty under the right conditions (determined by both spatial configuration and individual criterion). This is in contrast to a recent study by Braasch (2002), who presented similar stimuli to subjects in order to examine and model localization in the presence of a distracter. Braasch presented concurrent 200-ms broadband noises separated in azimuth in the frontal region (analogous to our 0° reference location), and reported that in the majority of presentations, subjects perceived only a *single* auditory image. This is surprising because in that study the separation angles ranged from 15° to 90°, and in the present study subjects could fully resolve stimuli at the maximum separation of 21°. As the stimuli were identical apart from a small difference in their duration, the different responses of subjects are most likely related to the different tasks which they were asked to perform. Whereas our subjects were asked to report the number of source locations they perceived, the primary task of Braasch's subjects was to give an estimate of the location of the auditory image (if there was only one) or the most lateral location (if there was more than one image). Perhaps the fact that our subjects were asked to focus exclusively on the number of sources allowed them to adopt a more sensitive criterion to this parameter.

It is important to consider exactly what cues the subjects in the present study may have used in their assessment of the number of sources in a stimulus. It is clear that the perception of more than one source was a binaural effect; noise pairs separated along the vertical midline remained perceptually fused. Taken together, the results suggested that the presence of two different ongoing ITDs in the presented signal was responsible for separation. However, there are at least two possible ways in which this ITD cue might be used by the auditory system: (1) the two ITDs might be explicitly extracted from the composite signal, or (2) the presence of

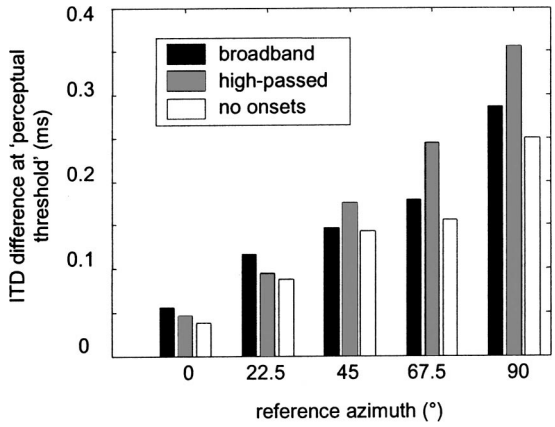


FIG. 6. Average ITD difference at perceptual threshold for experiments 1, 2a and 2b as a function of reference location. Values are pooled across subjects and direction of separation (see text for further details).

two ITDs in the mixed signal may be inferred from the neural representation. These two possibilities are discussed below.

B. A consideration of ITD sensitivity

If subject responses were driven by an ability to clearly perceive the two component ITDs, then it can be reasoned that the pattern of responses observed should bear some relation to ITD sensitivity with a single source. This idea is examined by considering the pattern of responses from experiments 1, 2a, and 2b which spanned the five reference locations.

Perceptual thresholds were estimated from the psychophysical curves for all subjects in these three conditions. Perceptual threshold was defined as the separation value whereby two source locations were reported to be perceived in 75% of cases. As ''positive'' and ''negative'' separation values were tested (corresponding to left and right, or up and down) an upper and lower threshold was obtained separately from the two halves of each curve. It was not possible to obtain a value in some cases, where performance did not reach the 75% level within the range of testing. For each case where a threshold value could be obtained, it was possible to express the angle in terms of the *ITD difference* between the two sources. Figure 6 shows the average ITD difference at perceptual threshold for each of the five reference locations. This value is pooled across the four subjects and the four directions of separation (left, right, down, up) for cases where threshold could be obtained.

It can be seen that, on average, listeners in the present experiment required about 50-μs difference in ITD to consistently perceive two sources when they were presented concurrently around the frontal reference position (0° azimuth). Klumpp and Eady (1956) made measurements of ITD thresholds for listeners using headphones. They found that for a broadband noise with a reference ITD of 0 μs, the threshold ITD for determining whether a successive stimulus was to the left or right was 10 μs. This value is substantially smaller than that for the present experiment, suggesting that, if ITD is the dominant cue in our task, it is more difficult to extract from a *concurrent* presentation of stimuli than from a sequential one. According to Fig. 6, the ITD difference required for separation increased as the reference location became more lateral. A similar effect has been reported for sequentially presented stimuli presented over headphones, where just-noticeable changes in ITD increase with baseline ITD. Klumpp and Eady (1956) used a band-pass filtered noise stimulus (150–1700 Hz) and found that the threshold for a 0-μs ITD stimulus was 9 μs but increased to 29 and 50 μs for ITDs of 430 and 790 μs, respectively. Other studies have shown similar effects for high- and low-frequency transients (Hafter and DeMaio, 1975) and 500-Hz tones (Hershkowitz and Durlach, 1969; Domnitz and Colburn, 1977).

The data in Fig. 6 also indicate that ITD thresholds are higher on average in the high-pass condition for the more lateral locations (45°, 67.5°, and 90° azimuth), confirming the observations made in experiment 2a. This finding is consistent with the study of Hafter and DeMaio (1975) where it was found that just-noticeable ITD differences were larger for high-frequency (3–4 kHz) clicks compared to low-frequency (0.1–2 kHz) clicks. These authors also reported that this disparity increased with increasing baseline ITD, and for the largest ITD examined (500 μs) ITD discrimination performance using high-frequency clicks was unmeasurable in one subject.

Thus, although there are substantial magnitude differences, it seems that there is some correspondence between the concurrent ITD thresholds estimated from the present data and reported sequential ITD discrimination thresholds. In both situations there is an increase in threshold with increasing reference ITD. In addition, while ITD in high-frequency regions is a useable cue, a low-frequency component improves performance at large values of ITD.

C. A consideration of ITD extraction from a mixed signal

Although the analysis presented above is consistent with the idea that ITDs are useful cues for the separation of broadband noise sources, one must consider by what mechanisms the auditory system might recover these cues from the summed stimuli it receives. Preliminary computational modeling using biologically plausible elements indicates that a running cross-correlation of the time-domain inputs to the left and right ears may provide an adequate explanation of these data (e.g., Best *et al.*, 2002). The emergence of two peaks in the cross-correlation function suggests that the auditory system may be able to extract individual ITD estimates for each of the sources.

However, an important effect of two independent sources mixing is decorrelation of the signals at the two ears. It is possible that this decorrelation is detected and used by subjects as an indication that there is more than one source. In terms of a cross-correlation mechanism, it may be that performance related less to the presence of two ITD peaks and more to the *lack of a strong peak*. Certainly it is clear from the literature that listeners are sensitive to interaural correlation (Pollack and Trittipoe, 1959a,b; Gabriel and Colburn, 1981) and that decorrelation is related to the perception of image width (e.g., Blauert and Lindemann, 1986). It is hoped that further experimentation and more sophisticated

TABLE I. Summary of localization data for the four subjects in the free-field and in virtual auditory space. Listed are spherical correlation coefficients and the percentage of trials in which a cone-of-confusion error occurred (in five tests each of 76 locations) under the two conditions.

Subject	Condition	SCC	% COC errors
1	Free-field	0.92	1.8
	VAS	0.85	3.7
2	Free-field	0.92	1.3
	VAS	0.92	2.1
3	Free-field	0.93	1.1
	VAS	0.90	2.9
4	Free-field	0.91	0.3
	VAS	0.91	1.8

modeling will illuminate this important issue of how well the ITDs of concurrent stimuli are preserved in the auditory system.

It is important also to note that the data in this study were collected under anechoic conditions, and the acoustics of real-world listening conditions may have some effect on the ability of subjects to discriminate one source from two. Certainly the presence of echoes could be expected to confuse the estimate of source numbers (by increasing the number of peaks and/or the interaural decorrelation). On the other hand, it has been shown that the impact of reverberation on the localization of a single source is minimal in some cases (Hartmann, 1983; Shinn-Cunningham, 2000).

VII. CONCLUSIONS

Spatial resolution with concurrent sources was examined systematically with a focus on the role of the different localization cues. The study was unique in that it (a) employed broadband stimuli that were distinguishable *only* on the basis of location, (b) focused on the detection of multiple sources in an acoustic stimulus rather than on issues of identification and/or localization, and (c) examined stimuli distributed both horizontally *and* vertically. It was shown that a concurrent pair of broadband white noises could be resolved if a sufficiently large binaural difference was present. In particular, differences in ongoing ITDs were shown to be robust cues for separation, including those in the envelopes of high-frequency channels (above 2 kHz).

APPENDIX: LOCALIZATION PERFORMANCE IN THE FREE-FIELD AND VIRTUAL AUDITORY SPACE

Prior to experimentation, the ability of the four subjects to localize broadband noise bursts in the free-field and in virtual auditory space (VAS) was assessed. Each subject underwent five localization tests (described in Sec. II B) in each condition, and the values in Table I summarize the data. Target and response locations were analyzed in terms of the lateral-polar angle coordinate system, and cone-of-confusion (COC) errors were defined as trials in which the target and response polar angle differed by more than 90°. The occurrence of this type of error is expressed as a percentage of the total number of trials. The remaining data were assessed statistically using the spherical correlation coefficient (SCC), which provides a measure of the correspondence between the perceived and actual directions on the sphere (1=perfect correlation; 0=no correlation). See Carlile *et al.* (1997) for more detail.

Batteau, D. W. (**1967**). ''The role of the pinna in human localization,'' Proc. R. Soc. London, Ser. B **158**, 158–180.

Bernstein, L. R., and Trahiotis, C. (**1982**). ''Detection of interaural delay in high-frequency noise,'' J. Acoust. Soc. Am. **71**(1), 147–152.

Best, V., van Schaik, A., and Carlile, S. (**2002**). ''The perception of multiple broadband noise sources presented concurrently in virtual auditory space,'' in *Proc. Audio Engineering Society 112th Convention*, Munich, Germany, paper 5549.

Blauert, J. (**1983**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge).

Blauert, J., and Lindemann, W. (**1986**). ''Spatial mapping of intracranial auditory events for various degrees of interaural coherence,'' J. Acoust. Soc. Am. **79**, 806–813.

Braasch, J. (**2002**). ''Localization in the presence of a distracter and reverberation in the frontal horizontal plane: II. Model algorithms,'' Acust. Acta Acust. **88**, 956–969.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge).

Bregman, A. S., and Campbell, J. (**1971**). ''Primary auditory stream segregation and perception of order in rapid sequences of tones,'' J. Exp. Psychol. **89**, 244–249.

Buell, T. N., Trahiotis, C., and Bernstein, L. R. (**1991**). ''Lateralization of low-frequency tones: Relative potency of gating and ongoing interaural delays,'' J. Acoust. Soc. Am. **90**, 3077–3085.

Butler, R. A. (**1986**). ''The bandwidth effect on monaural and binaural localization,'' Hear. Res. **21**, 67–73.

Butler, R. L., and Humanski, R. A. (**1992**). ''Localization of sound in the vertical plane with and without high-frequency spectral cues,'' Percept. Psychophys. **51**(2), 182–186.

Butler, R. A., Humanski, R. A., and Musicant, A. D. (**1990**). ''Binaural and monaural localization of sound in two-dimensional space,'' Perception **19**, 241–256.

Carlile, S. (ed.) (**1996**). *Virtual Auditory Space: Generation and Applications* (Landes, Austin).

Carlile, S., and Delaney, S. (**1999**). ''The localization of spectrally restricted sounds by human listeners,'' Hear. Res. **128**, 175–189.

Carlile, S., and Leung, J. (**2001**). ''Rendering sound sources in high fidelity virtual auditory space: Some spatial sampling and psychophysical factors,'' in *Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents and Virtual Reality*, edited by M. Smith, G. Salvendy, D. Harris, and R. Koubek (Erlbaum, Hillsdale, NJ), pp. 599–603.

Carlile, S., Jin, C., and van Raad, V. (**2000**). ''Continuous virtual auditory space using HRTF interpolation: Acoustic and psychophysical errors,'' *Proc. First IEEE Pacific-Rim Conf. Multimedia*, Sydney, Australia, pp. 220–223.

Carlile, S., Leong, P., and Hyams, S. (**1997**). ''The nature and distribution of errors in sound localization by human listeners,'' Hear. Res. **114**, 179–196.

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.

Divenyi, P. L., and Oliver, S. K. (**1989**). ''Resolution of steady-state sounds in simulated auditory space,'' J. Acoust. Soc. Am. **85**, 2042–2052.

Domnitz, R. H., and Colburn, H. S. (**1977**). ''Lateral position and interaural discrimination,'' J. Acoust. Soc. Am. **61**, 1586–1598.

Durlach, N. I., and Colburn, H. S. (**1978**). ''Binaural Phenomena,'' in *The Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman (Academic, New York).

Gabriel, K. J., and Colburn, H. S. (**1981**). ''Interaural correlation discrimination: I. Bandwidth and level dependence,'' J. Acoust. Soc. Am. **69**, 1394–1401.

Gaik, W. (**1993**). ''Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling,'' J. Acoust. Soc. Am. **94**, 98–110.

Gardner, M. B., and Gardner, R. S. (**1973**). ''Problem of localization in the median plane: Effect of pinnae cavity occlusion,'' J. Acoust. Soc. Am. **53**, 400–408.

Hafter, E. R., and De Maio, J. (**1975**). ''Difference thresholds for interaural delay,'' J. Acoust. Soc. Am. **57**, 181–187.

Hartmann, W. M. (**1983**). ''Localization of sound in rooms,'' J. Acoust. Soc. Am. **74**, 1380–1391.

Henning, G. B. (**1974**). ''Detectability of interaural delay in high-frequency complex waveforms,'' J. Acoust. Soc. Am. **55**, 84–90.
Henning, G. B. (**1980**). ''Some observations on the lateralization of complex waveforms,'' J. Acoust. Soc. Am. **68**, 446–454.
Hershkowitz, R. M., and Durlach, N. I. (**1969**). ''Interaural time and amplitude jnds for a 500-Hz tone,'' J. Acoust. Soc. Am. **46**, 1464–1467.
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.
Kunov, H., and Abel, S. M. (**1981**). ''Effects of rise/decay time on the lateralization of interaurally delayed 1-kHz tones,'' J. Acoust. Soc. Am. **69**, 769–773.
McFadden, D., and Pasanen, E. G. (**1976**). ''Lateralization at high frequencies based on interaural time differences,'' J. Acoust. Soc. Am. **59**, 634–639.
Middlebrooks, J. C. (**1992**). ''Narrow-band sound localization related to external ear acoustics,'' J. Acoust. Soc. Am. **92**, 2607–2624.
Middlebrooks, J. C., and Green, D. M. (**1991**). ''Sound localization by human listeners,'' Annu. Rev. Psychol. **42**, 135–159.
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.
Mills, A. W. (**1972**). ''Auditory Localization,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), pp. 303–348.
Moller, H., Sorensen, M. F., Hammershoi, D., and Jensen, C. B. (**1995**). ''Head-related transfer functions of human subjects,'' J. Audio Eng. Soc. **43**(5), 300–321.
Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing* (Academic, London).
Noble, W., Byrne, D., and Ter-Host, K. (**1997**). ''Auditory localization, detection of spatial separateness, and speech hearing in noise by hearing impaired listeners,'' J. Acoust. Soc. Am. **102**, 2343–2352.
Perrott, D. R. (**1984**). ''Concurrent minimum audible angle: A re-examination of the concept of auditory spatial acuity,'' J. Acoust. Soc. Am. **75**, 1201–1206.
Perrott, D. R., and Saberi, K. (**1990**). ''Minimum audible angle thresholds for sources varying in both elevation and azimuth,'' J. Acoust. Soc. Am. **87**, 1728–1731.
Pickles, J. O. (**1988**). *An Introduction to the Physiology of Hearing* (Academic, London).
Pollack, I., and Trittipoe, W. J. (**1959a**). ''Binaural listening and interaural noise cross correlation,'' J. Acoust. Soc. Am. **31**, 1250–1252.
Pollack, I., and Trittipoe, W. J. (**1959b**). ''Interaural noise correlations: Examination of variables,'' J. Acoust. Soc. Am. **31**, 1616–1618.
Pralong, D., and Carlile, S. (**1996**). ''The role of individualized headphone calibration for the generation of high fidelity virtual auditory space,'' J. Acoust. Soc. Am. **100**, 3785–3793.
Shinn-Cunningham, B. G. (**2000**). ''Learning reverberation: Considerations for spatial auditory displays,'' in *Proc. Int. Conf. Auditory Display*, Atlanta, GA, pp. 126–134.
Stellmack, M. A. (**1994**). ''The reduction of binaural interference by the temporal nonoverlap of components,'' J. Acoust. Soc. Am. **96**, 1465–1470.
Tobias, J. V., and Schubert, E. D. (**1959**). ''Effective onset duration of auditory stimuli,'' J. Acoust. Soc. Am. **31**, 1595–1605.
Yost, W. A. (**1976**). ''Lateralization of repeated filtered transients,'' J. Acoust. Soc. Am. **60**, 178–181.
Zahorik, P. (**2000**). ''Limitations in using Golay codes for head-related transfer function measurement,'' J. Acoust. Soc. Am. **107**, 1793–1796.
Zhou, B., Green, D. M., and Middlebrooks, J. C. (**1992**). ''Characterization of external ear impulse responses using Golay codes,'' J. Acoust. Soc. Am. **92**, 1169–1171.

A model of acoustic interspeaker variability based on the concept of formant–cavity affiliation

Lian Apostol, Pascal Perrier,[a] and Gérard Bailly
Institut de la Communication Parlée, UMR CNRS 5009, INPG, Grenoble, France

(Received 10 December 2002; accepted for publication 20 October 2003)

A method is proposed to model the interspeaker variability of formant patterns for oral vowels. It is assumed that this variability originates in the differences existing among speakers in the respective lengths of their front and back vocal-tract cavities. In order to characterize, from the spectral description of the acoustic speech signal, these vocal-tract differences between speakers, each formant is interpreted, according to the concept of formant–cavity affiliation, as a resonance of a specific vocal-tract cavity. Its frequency can thus be directly related to the corresponding cavity length, and a transformation model can be proposed from a speaker A to a speaker B on the basis of the frequency ratios of the formants corresponding to the same resonances. In order to minimize the number of sounds to be recorded for each speaker in order to carry out this speaker transformation, the frequency ratios are exactly computed only for the three extreme cardinal vowels [i, a, u] and they are approximated for the remaining vowels through an interpolation function. The method is evaluated through its capacity to transform the $(F1,F2)$ formant patterns of eight oral vowels pronounced by five male speakers into the $(F1,F2)$ patterns of the corresponding vowels generated by an articulatory model of the vocal tract. The resulting formant patterns are compared to those provided by normalization techniques published in the literature. The proposed method is found to be efficient, but a number of limitations are also observed and discussed. These limitations can be associated with the formant–cavity affiliation model itself or with a possible influence of speaker-specific vocal-tract geometry in the cross-sectional direction, which the model might not have taken into account. [DOI: 10.1121/1.1631946]

PACS numbers: 43.70.Gr, 43.70.Bk [AL] Pages: 337–351

I. INTRODUCTION

For a long time, the study of variability in speech was mainly guided by speech recognition objectives (Klatt, 1986; Stevens, 1980; Perkell and Klatt, 1986). The aim was thus basically to quantify and to characterize variability in the temporal and spectral domains, not to reproduce it, but to eliminate it, in order to extract *the invariant*, the physical pattern associated with the linguistic input to be recovered. In other words, the aim was to ''normalize'' the speech signal among speakers. For vowels, the purpose of such a technique was to reduce, in the acoustic domain, the variance of the data measured for the same vowel on a number of speakers, in order to enhance the identification scores of the sounds in an automatic classification.

More recently (see in particular Story and Titze, 2002; Titze *et al.*, 1996, 1997; Wong *et al.*, 1996), special attention has been devoted to the generation of interspeaker variability with the aim of contributing to the development of multivoice and multispeaker speech synthesis systems. The present paper is in the line of these studies. Our aim is to propose a speaker transformation technique in the acoustic domain, based on an account of the correlates of acoustic variability in the domain of speaker-specific vocal-tract geometry, with the final objective of predicting the variability in the whole acoustic domain from a limited amount of speech samples for each speaker.

Our basic hypothesis is that the largest amount of interspeaker variability of the formant patterns arises from differences among speakers in the respective lengths of their back and front vocal-tract cavities. These geometrical differences (see, e.g., Bothorel *et al.*, 1986, as an illustration) are due to intrinsic morphological properties of the vocal tract, such as its length (as illustrated by Subjects 3 and 4 in Bothorel *et al.*, 1986), or its distribution between the palatal and the pharyngeal regions (as illustrated by Subjects 1 and 3 in Bothorel *et al.*, 1986), but these differences can also be due to speaker-specific articulatory strategies involved in the positioning and in the displacement of the tongue in the vocal tract (for an example see the [i] production of Subjects 2 and 3 in Bothorel *et al.*, 1986, pp. 16–17). Consequently, in order to transform for each vowel production the formant pattern of speaker A into the formant pattern of speaker B, our approach consists of elaborating a general model for the changes in front and back cavity lengths between these speakers. This model is based on the computation, from the formant frequencies, of back and front cavity length ratios. To do this, it is proposed to apply a basic principle of the acoustic theory of vowel production, namely the formant–cavity affiliation principle. This principle suggests that for a given vowel each formant can be associated, more or less tightly, with a specific cavity of the vocal tract (Chiba and Kajiyama, 1941; Dunn, 1950; Fant, 1960; Stevens, 1972, 1989, 1998), and is therefore essentially determined by the geometry of this cavity.

In this paper, after a short résumé of the main aspects of

[a]Electronic mail: perrier@icp.inpg.fr

this theory and, particularly, of its interpretation in terms of articulatory-to-acoustic relations in vowel production, a method based on these principles and called the *resonance-based method (RBM)* will be presented. A quantitative evaluation of the *RBM* will then be proposed by assessing its capacity to transform the $(F1,F2)$ patterns of five male speakers into the formant patterns of a reference articulatory model of the vocal tract. In order to assess our hypotheses carefully, the corpus we have used consisted of a reduced number of well-controlled logatoms, and, for the purposes of this paper, the variability in connected speech has not been addressed. On the basis of this evaluation, the intrinsic strengths and limits of the concept of affiliation between formants and cavities will be discussed.

II. FORMANTS AND VOCAL-TRACT RESONANCES

A. The formant–vocal-tract cavity affiliation (Fant, 1960)

Fant (1960) has shown that it is possible to obtain fairly good predictions of the formant patterns characterizing oral vowels by using a simple modeling of the vocal tract, consisting of only four tubes. Such a simplification allowed the formants to be specifically interpreted as resonance frequencies of the different cavities of the vocal tract.[1] In this perspective, a basic and efficient tool consists of the well-known nomograms presented by Fant (Fant, 1960, p. 76), which show the variations of the first five formants, when the vocal-tract constriction is shifted from the glottis to the lips: ''*if an advance of the tongue causes a resonance frequency to rise, it can be concluded that the resonance is mainly influenced by a cavity of decreasing length*'' (Fant, 1960, p. 75). This interdependence of resonance and the vocal-tract cavity is all the more evident if the constriction area is small, reducing the acoustical coupling between vocal-tract cavities.

In a vocal tract where cavities are essentially uncoupled, the following resonance modes can be observed (see Fant, 1960, for details):

(i) A half-wavelength resonance mode: its characteristic frequencies are given by the formula

$$R_{\lambda/2}=n\frac{c}{2L}, \tag{1}$$

where c is the sound velocity in the air, L is the length of the considered cavity, and $n \in \mathbf{N}$.

(ii) A quarter-wavelength resonance mode: its characteristic frequencies are given by the formula

$$R_{\lambda/4}=(2n-1)\frac{c}{4L}, \tag{2}$$

with the same notations as above.

(iii) A Helmholtz resonance mode: its resonance frequency is given by the relation

$$R_H=\frac{c}{2\pi}\sqrt{\frac{A}{l\cdot V}}, \tag{3}$$

in which V is the volume of the cavity, and A and l are, respectively, the area and the length of the resonator's ''neck.''

When vocal-tract cavities are acoustically coupled, the relations between geometry and formants do not strictly apply: the more the coupling between cavities, the less the model is valid. In addition, a clear affiliation of formants and cavities happens to be very difficult when the constriction is located in the so-called *focal* regions of the vocal tract. For a vocal tract having its constriction in one of these *focal* regions, the resonance frequencies of the uncoupled cavities have very similar values (see Badin *et al.*, 1990). The size of these regions in the anterior–posterior direction increases with the acoustical coupling between cavities. This phenomenon adds to the difficulty of finding reliable affiliations in the case of a significant acoustical coupling. Note, however, that studying formant variations when the position of the constriction is moved step by step through a focal region of the vocal tract along the back/front direction can help to reduce the uncertainty about affiliations in such a region (see below, Sec. III D).

Having established these two basic alternatives, we should note that reliable formant–cavity affiliations can be hypothesized for a relatively weak coupling between cavities, and outside of the focal regions (Fant, 1960; Mrayati and Carré, 1976; Badin *et al.*, 1990). Fant (1960) and Badin *et al.* (1990) determined these mappings by localizing each vowel on Fant's nomograms so that its ''typical'' formant pattern is produced with plausible constriction position and lip area.

In such conditions, it becomes possible to infer global morphological differences between speakers by simply comparing their respective formants and interpreting them as specific vocal-tract resonances, while assuming that the coupling between cavities for a given vowel is essentially constant among speakers. This is the basic principle of the ''*resonance-based method*'' (called the RBM) that we have formulated to account for interspeaker variability.

B. The resonance-based method

It is known that formant values are influenced by the 3D geometry of the vocal tract. A Helmholtz resonance depends [see Eq. (3)] on the volume of the resonator's ''body,'' as well as on the length and on the cross-sectional area of the resonator's ''neck.'' The half- and quarter-wavelength resonances, which, in theory, depend only on the length (measured parallel to the direction of the air flow) of the associated cavity, are also influenced by the coupling between cavities that depends on the length and on the cross-sectional area of the constriction. Therefore, both differences in cavity lengths and in cavity cross-sectional areas are *a priori* likely to generate interspeaker differences in the acoustic domain. Nevertheless, in the continuity of classical models of interspeaker variability in vowel production published in the literature (e.g., Nordström and Lindblom, 1975; Wakita, 1977; Fant, 1975), for the purposes of this paper we are going to assume that interspeaker differences in cavity lengths are the major factor of interspeaker variability in the formant domain. For formants associated with half-wavelength or quarter-wavelength resonances, this hypothesis means that the influence of interspeaker differences in constriction size, and then in acoustic coupling between cavities, is signifi-

cantly less important than the influence of differences in cavity lengths. This assumption is consistent with the nomograms proposed by Fant (1960) with the four-tube model for two different sections of the constriction tube (see pp. 76–77) and with the horn-shape model for three different minimum constriction areas of the vocal tract (see p. 84). As concerns formants associated with a Helmholtz resonator, the above hypothesis means that the ratio of the body and neck cross-sectional areas has significantly less impact than the ratio between the lengths of these cavities. To our knowledge, this assumption has never been experimentally demonstrated. The present study will contribute to quantitatively evaluate its validity.

According to Fant's nomograms, the number of possible affiliations for the first three formants of an oral vowel is very limited: they are either a quarter-wavelength or a half-wavelength resonance of the front or the back cavity or a Helmholtz resonator of the set ''back cavity+constriction,'' or of the set ''front cavity+lips'' in the case of a rounded vowel (Fant, 1960). Thus, in the framework of this acoustic theory of vowel production, our assumption that interspeaker differences in the area of the vocal-tract constriction can be neglected implies that the variability of the first three formants observed between two speakers *X* and *Y* is explained by the variabilities of the back and front cavity lengths. Therefore, for each vowel v, two length ratios are taken into consideration, one for the front cavity *f*, and the other one for the back cavity *b*

$$\alpha_c[v]=\frac{L_{c,X}[v]}{L_{c,Y}[v]}, \quad c\in\{f,b\}, \tag{4}$$

where $L_{c,X}[v]$ et $L_{c,Y}[v]$ are the lengths of the *c* cavity for speaker *X* and *Y* respectively.

To infer these ratios from the acoustic signal, the relations between resonance frequencies and cavity lengths are used (see equations 1–3). Thus, for half-wavelength as well as for quarter-wavelength resonances, the ratio is expressed as

$$\alpha_c[v]=\frac{L_{c,X}[v]}{L_{c,Y}[v]}=\frac{R_{c,Y}}{R_{c,X}}, \quad c\in\{f,b\}, \tag{5}$$

where $R_{c,X}[v]$ and $R_{c,Y}[v]$ are the lowest half- or quarter-wavelength resonances of the *c* cavity for speaker *X* and *Y*, respectively. For the Helmholtz mode, the length L_c of the resonator's body is not the unique factor determining the resonance frequency R_H. However, according to our hypotheses that interspeaker differences in vocal-tract cross-sectional area and in cavities coupling can be neglected, it can be assumed that the ''shape factor'' $\sqrt{A/l}$ of the resonator's neck is nearly constant, and that differences in volume *V* of the resonator body are mainly due to differences of its length L_c. Under these conditions, the corresponding length ratio can be calculated according to the formula

$$\alpha_c[v]=\frac{L_{c,X}[v]}{L_{c,Y}[v]}\approx\frac{R^2_{H,c,Y}}{R^2_{H,c,X}}, \quad c\in\{f,b\}. \tag{6}$$

In this study, the length ratios will be calculated as formant frequency ratios on the basis of hypotheses concerning the affiliations between formants and cavities. Testing the RBM will allow us to test the validity of these hypotheses.

III. IMPLEMENTATION OF THE RBM FOR SPEAKER TRANSFORMATION IN VOWEL PRODUCTION

A. Selected basis vowels

The RBM was used to elaborate a speaker transformation procedure in vowel production. The aim of this procedure is to be able to generate every vowel sequence of a ''target'' speaker *X* from the recording of the same sequence pronounced by ''source'' speaker *Y*. To do this, the basic idea is to infer general transformation rules from a limited number of vowel sounds recorded both for speaker *X* and *Y*. The choice of these specific sounds is crucial, since they have to carry enough information about the speech articulation of each speaker, so that a generalization concerning the whole vowel space could be possible through our method. For this reason, we decided to use the extreme cardinal vowels [i, u, a]. Indeed, since they correspond in theory to the most extreme speaker-specific articulatory configurations, they effectively cover to a large extent the articulatory space used in vowel production: from [i] to [u], the tongue has a high position in the vocal tract and the lingual constriction moves along the sagittal palatal contour, from the most anterior position to the most posterior one; from [u] to [a], the tongue moves down in the vocal tract from the highest to the lowest position. We'll call these vowels *basis vowels* of the RBM in the rest of this text.

For these *basis vowels* speaker-specific information about the articulatory strategy was extracted through a quantitative comparison with a reference articulatory model of the vocal tract, whose acoustic and articulatory characteristics (i.e., the place of articulation, the area function, and the formants), are precisely known, and whose formant affiliations can be properly determined among the different possibilities suggested by Fant's nomograms.

B. Characteristics of the reference articulatory model of the vocal tract

We used a statistical articulatory model, *Bergame*, developed at the *Institut de la Communication Parlée* in Grenoble on a French speaker (Beautemps *et al.*, 1996) following the method proposed by Maeda (1990). Beautemps *et al.* (1996) performed a ''*statistical analysis of midsagittal vocal tract profiles derived from cineradiographic pictures, recorded in synchrony with video pictures of front views of the lips and with the speech signal.*'' This model generates an area function of the vocal tract from seven articulatory parameters. The associated formant patterns were obtained with the harmonic vocal tract model developed by Badin and Fant (1984).[2]

1. Reference vowels

Reference articulatory configurations were generated with the reference articulatory model of the vocal tract for eight French vowels [i, e, ɛ, a, y, u, o, ɔ]. To do this, two constraints were respected. The first constraint consisted in maintaining inside the whole vowel system realistic articula-

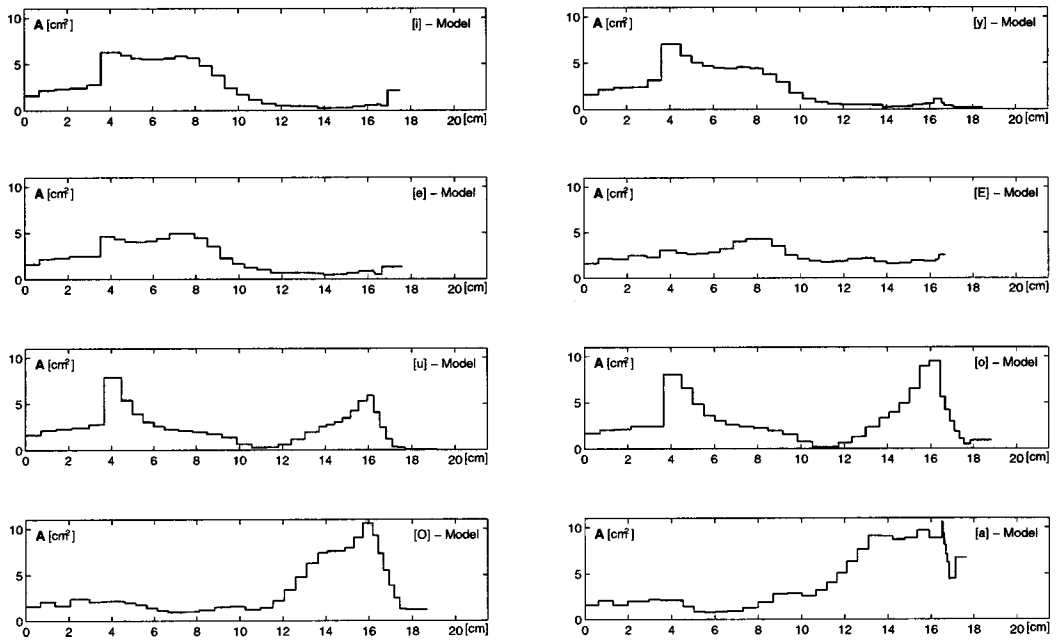


FIG. 1. Area functions obtained for the reference speaker; from the top and from the left to the right: [i, y, e, ɛ, u, o, ɔ, a].

tory positions (Straka, 1965; Abry and Boë, 1986; Bothorel *et al.*, 1986; Majid *et al.*, 1987; Boë *et al.*, 1992): jaw position becomes progressively lower from [i] to [a], passing through [e] and [ɛ]; the lips are more rounded for [y] and [u] than for [o] and [ɔ]; the constriction position in the vocal tract moves back from [i] to [a], passing through [e] and [ɛ], and goes down along the pharynx from [u] to [a], passing through [o] and [ɔ]. The second constraint consisted of ensuring that the formant values of the synthetic reference vowels were close to the ones measured on the corresponding tokens of the French speaker used to develop *Bergame* (Beautemps *et al.*, 1996).

Thus, a set of commands to the articulatory model was obtained for each reference vowel by an acoustic-to-articulatory inversion.[3] A sagittal function, an area function, and the resulting formants were then calculated for each configuration. The eight area functions are given in Fig. 1.

2. Formant–cavity affiliations for the reference vowels

In order to establish the formant–cavity affiliations for the eight reference vowels, our approach was inspired by the vocal-tract sensitivity functions proposed by Fant and Pauli (1974). It consists first of increasing by 10% the cross-sectional area for each elementary section of the area function separately, and, second, in calculating the induced relative formant variations, which we called *sensitivities*. A study carried out on simple tubes, the theoretical resonances of which are known, provided a set of reference patterns for the sensitivities associated with the main resonance modes suggested by Fant's nomograms (1960).

a. Reference sensitivity patterns. Four simple shaped acoustic tubes were chosen for this preliminary study. Each one was obtained by the concatenation of several elementary tubes 0.425 cm in length, so as to reach a total length of 17 cm, which is comparable with the mean length of a male vocal tract (Stevens, 1998). The first tube is closed at one end and has a constant cross-sectional area of 4 cm^2. The first five formants are, in this case, odd multiples of the quarter-wavelength mode of the whole tube ($\lambda/4$, $3\lambda/4$, $5\lambda/4$, etc.). The second tube is similar to the first one, except that it is closed at both ends, its last three sections having a very small cross-sectional area (0.3 cm^2). The first formant of this configuration is therefore a Helmholtz resonance, whereas the next ones are multiples of the half-wavelength mode ($\lambda/2$, λ, $3\lambda/2$, etc.) of the large cavity. The sensitivities associated with the first five formants of these tubes are given in Fig. 2. Two more configurations were obtained by coupling two tubes, which are similar to those described above, with a small tube having a small cross-sectional area. These models are rough representations of a pharyngeal vowel with a large lip opening, and of a rounded velar vowel (see Fig. 3). In the first case, note that the Helmholtz resonance (given by $F1$) concerns only the set "back cavity+constriction," whereas $F2$, $F3$, and $F5$ are the $\lambda/4$, $3\lambda/4$, and $5\lambda/4$ modes of the front cavity, respectively. $F4$ is the half-wavelength resonance of the open–open tube that represents the constriction. In the second case formant sensitivities are those of a couple of Helmholtz resonators.

b. Establishing the formant–cavity affiliations for the reference vowels. The computation of the formant sensitivity curves for the eight area functions of *Bergame* representing the reference vowels permitted for each of them the formulation of the most plausible formant–cavity affiliations. For that, the affiliation was determined first by looking for each formant at the location of the largest sensitivity. Then, the most plausible nature of the resonance (Helmholtz, quarter-wavelength, or half-wavelength) was proposed by comparing the area function and the sensitivity curves with those of the reference patterns. As an example, Fig. 4 shows the formant sensitivities for the vowels [a] (left panel) and [u] (right panel), which are to be compared with the patterns given in

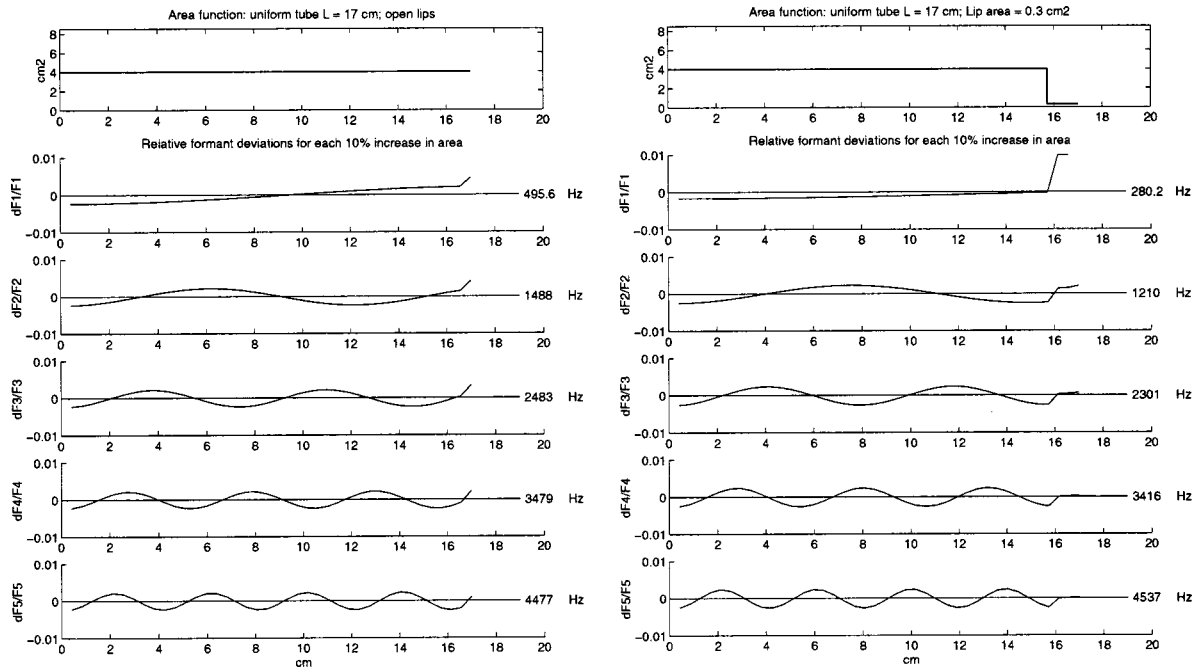


FIG. 2. Formant sensitivities to local area perturbations. Left panel: a closed–open uniform tube. Right panel: a closed–closed uniform tube.

Fig. 3. The affiliations thus established together with the corresponding first four formant values are summarized in Table I.

For all reference vowels but one, the interpretation of the computed sensitivities on the basis of the reference sensitivity patterns led unambiguously to formant–cavity affiliations compatible with a schematization of the area functions with a tube model (see Fig. 3). Vowel [ɛ] is, however, a particular case. It corresponds to an open vocal-tract configuration, without any true constriction separating clearly the vocal tract into two distinct cavities. As a result, the sensitivity curves of this reference vowel (see Fig. 5) are similar to those of the reference pattern depicted in the left panel of Fig. 2: formants are essentially affiliated with the whole vocal tract and they are odd multiples of the quarter-wavelength resonance of the whole vocal tract.

C. Interpolation between basis vowels

As explained above, in our method, the articulatory configurations of the ''target'' speaker *X* and of the ''source'' speaker Y are inferred from the formant patterns of the three

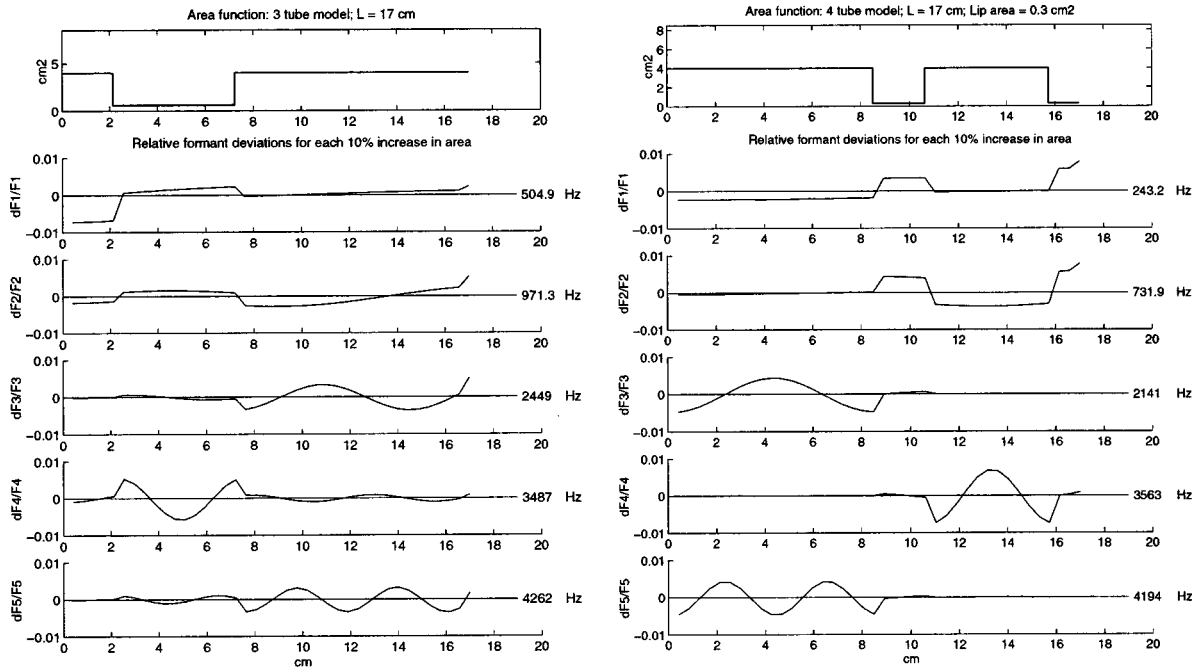


FIG. 3. Formant sensitivities to local area perturbations. Left panel: a tube model simulating a pharyngeal articulation with a large lip opening. Right panel: a tube model simulating a velar articulation with rounded lips.

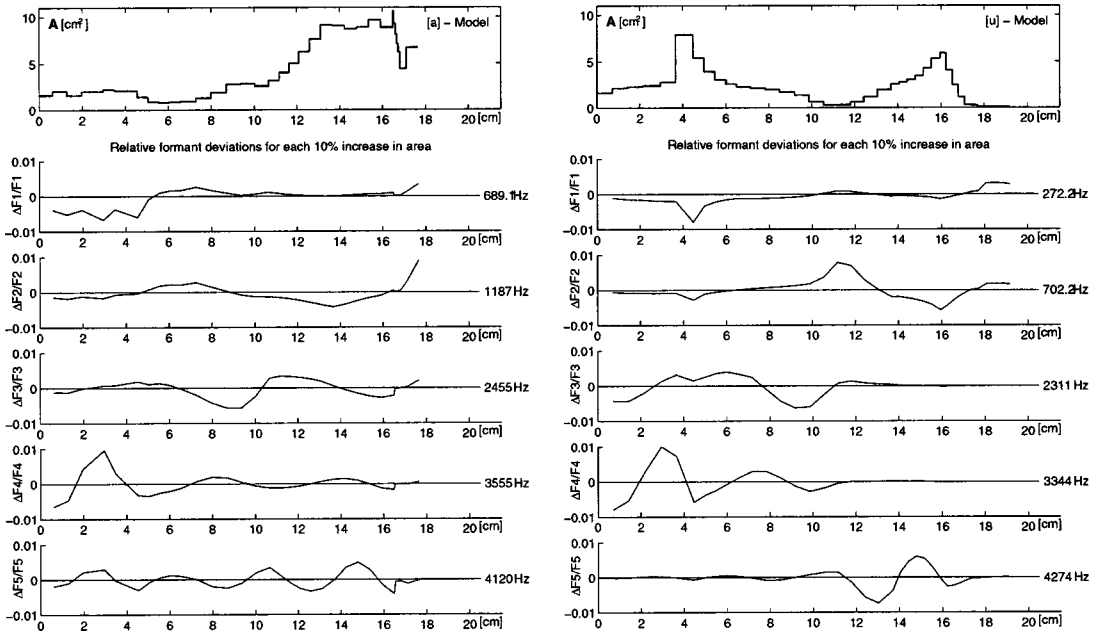


FIG. 4. Formant sensitivities for the reference vowels [a] (left panel) and [u] (right panel) generated with the reference articulatory model of the vocal tract (for comparison see, respectively, the left and right panels of Fig. 3).

basis vowels [a], [i], [u], which give a good account of the speaker's maximal vowel space. These basis vowels are systematically recorded both for the source and for the target speaker. Hence, for the front and for the back cavity, the ratio of the length of each cavity between speakers (α_b for the back cavity and α_f for the front cavity) can be found by the direct computation of the appropriate formant ratios, once the association between formants and cavity resonances is determined (cf. Sec. II B). For the other vowels, which are only recorded for the source speaker, the direct calculation is not possible. Hence, interpolation functions were used to approximate the interspeaker length ratios for these other vowels. For the vowels articulated in the palatal region, like [y], [e], and [ɛ], the coefficients were obtained using a logarithmic interpolation between the ratios of the *basis vowels* [i] and [u]. In a similar way, another logarithmic interpolation between [u] and [a] gave the ratios of the vowels [o] and [ɔ]. The interpolation functions depend on the constriction position measured on the reference vowels (cf. Sec. III A 1).

Thus, for [y], [e], and [ɛ], the ratios were calculated according to the formula

TABLE I. Formant-resonance associations for the vowel prototypes of the *Bergame* model. Legend: *H. back*=Helmohotz resonance of the set ''back cavity+constriction;'' *H. front*=Helmholtz resonance of the set ''front cavity+lips;'' for the other resonances: *back*=back cavity; *front*=front cavity.

Vowel	*F*1	*F*2	*F*3	*F*4
[i]	290	2069	2935	3669
	H. back	λ/2 back	λ/4 front	λ back
[y]	274	1766	2256	3264
	H. back	H. front	λ/2 back	λ back
[e]	349	1932	2641	3638
	H. back	λ/2 back	λ/4 front	λ back
[ɛ]	512	1702	2542	3796
	λ/4	3λ/4	5λ/4	7λ/4
	whole tract	whole tract	whole tract	whole tract
[u]	273	703	2311	3344
	H. back	H. front	λ/2 back	λ back
[o]	340	831	2316	3278
	H. back	H. front	λ/2 back	λ back
[ɔ]	522	911	2364	3370
	H. back	H. front	λ/2 front	λ/2 back
[a]	689	1187	2455	3555
	H. back	λ/4 front	3λ/4 front	λ/2 back

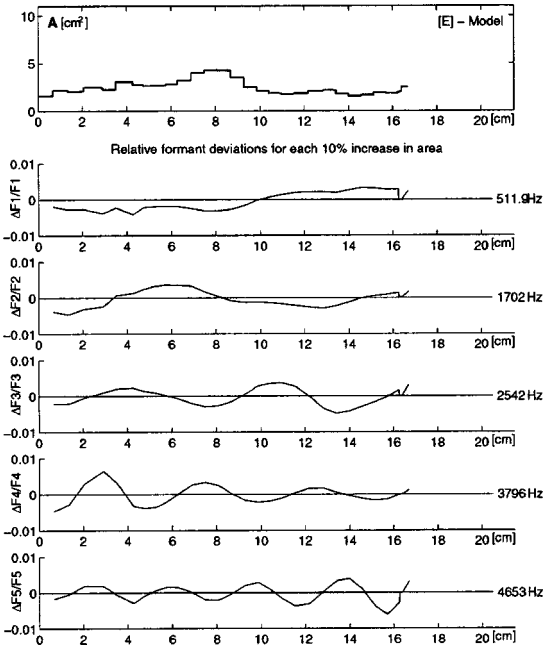


FIG. 5. Formant sensitivities for the reference vowel [ɛ] generated with the reference articulatory model of the vocal tract (for comparison, see Fig. 2, left panel).

$$\alpha_c[v]=\alpha_c[i]+(\alpha_c[u]-\alpha_c[i])\cdot\frac{\log\left(1+CO_c\dfrac{DC[v]-DC[i]}{DC[i]}\right)}{\log\left(1+CO_c\dfrac{DC[u]-DC[i]}{DC[i]}\right)}, \tag{7}$$

and for vowels [o] and [ɔ], α values were obtained from the equation

$$\alpha_c[v]=\alpha_c[a]+(\alpha_c[u]-\alpha_c[a])\cdot\frac{\log\left(1+CP_c\dfrac{DC[a]-DC[v]}{DC[a]}\right)}{\log\left(1+CP_c\dfrac{DC[a]-DC[u]}{DC[a]}\right)}, \tag{8}$$

where $c\in\{f,b\}$, and $DC[v]$ is the distance from the incisors to the center of the constriction for the reference vowel v; CO_c and CP_c are two coefficients that were experimentally determined for each cavity, in order to obtain a good level of efficiency in this procedure.

The calculation of the DC values necessitated the computation of the position of the constriction center for each reference vowel. To do this, the limits of the vocal-tract constriction were determined as the extremities of the vocal-tract's region within which an increase of the cross-sectional area of each elementary tube generates a variation of $F1$ compatible with a Helmholtz resonance: since in this case the constriction is the resonator's neck, $F1$ should increase when the cross-sectional area of the constriction increases (see the positive $\Delta F1/F1$ in the regions of the constriction in Figs. 3 and 4). The position of the constriction center was then determined as the abscissa separating the constriction zone in two parts of equal acoustic impedance. To calculate this abscissa, each elementary acoustic tube was replaced by its low-frequency electrical equivalent (cf. Fant, 1960, p. 28, Fig. 1.2-1). The DC values thus obtained are given in Table II.

D. The approach proposed to determine formant–cavity affiliations for real speakers

Obviously, since the RBM applies to speakers whose vocal-tract geometry is not known, the above approach based on sensitivities of the formants to local geometrical changes could not be used in order to find out the formant–cavity affiliations for real speakers. Therefore, the formant–resonance associations established for the reference vowels were used as initial patterns. However, these relations may not be valid for every speaker, in particular for the three basis vowels, and for vowel [y], because these vowels are located in focal regions of the vocal tract where small changes in vocal-tract geometry can cause changes in the affiliation pattern: this generates an uncertainty about the $F1-F2$ affiliation for [a] and [u]; and about $F2-F3$ or $F3-F4$ affiliation for [i] and [y], (cf. Badin *et al.*, 1990). This is why, in addition to the three *basis* vowels, vowel [y] was also recorded both for the source and for the target speaker, and the affiliation patterns of these four focal vowels were carefully analyzed for each speaker.

For vowel [u], the work of Savariaux *et al.* (1995) on 11 French native speakers showed that $F2$ represents in all cases the Helmholtz resonance of the set "front cavity + lips," whereas $F1$ counts for the Helmholtz resonance of the set "back cavity + constriction." Consequently, we adopted this affiliation pattern for this vowel.

As concerns the other "focal" vowels [i], [a], and [u], there is no such experimental evidence in the literature supporting one affiliation pattern more than another. This is why we recorded for each speaker a number of $V1-V2$ sequences, where $V1$ and $V2$ are either [i], [a], or [u], and analyzed the formant trajectories in these sequences. Indeed, while abrupt changes in formant frequencies can be observed during such sequences when the articulation location goes through a focal point, resonance frequencies vary smoothly and monotonically (Bailly, 1993). Consequently, knowing the articulatory changes involved in each $V1-V2$ sequence, and looking for the corresponding monotonous resonance variations, it was possible to make reliable assumptions about the formant–cavity affiliations.

For example, in an [iy] sequence, while large changes are observed in the lip shape, from spread (for [i]) to rounded (for [y]) lips, the back cavity undergoes basically no modification. Indeed, both the place of articulation and the size of the constriction are similar for the two vowels (cf. Bothorel *et al.*, 1986). Consequently, it is reasonable to consider that the resonance that varies the most during the sequence is affiliated with the front cavity. For [y] the lowest resonance frequency of the front cavity corresponds to $F2$. Therefore, following the variation of the resonance backwards across the sequence allows us to find for [i] which is the lowest formant affiliated to the same cavity. An illustration of the analysis is given in Fig. 6, which presents the formant track-

TABLE II. Incisors-to-center of the constriction distances (in cm) for the vowels of the reference articulatory model of the vocal tract.

Vowel	[i]	[y]	[e]	[ɛ]	[u]	[o]	[ɔ]	[a]
DC [cm]	2.4854	2.7776	2.6030	2.9374	6.0975	6.2873	9.3164	9.9971

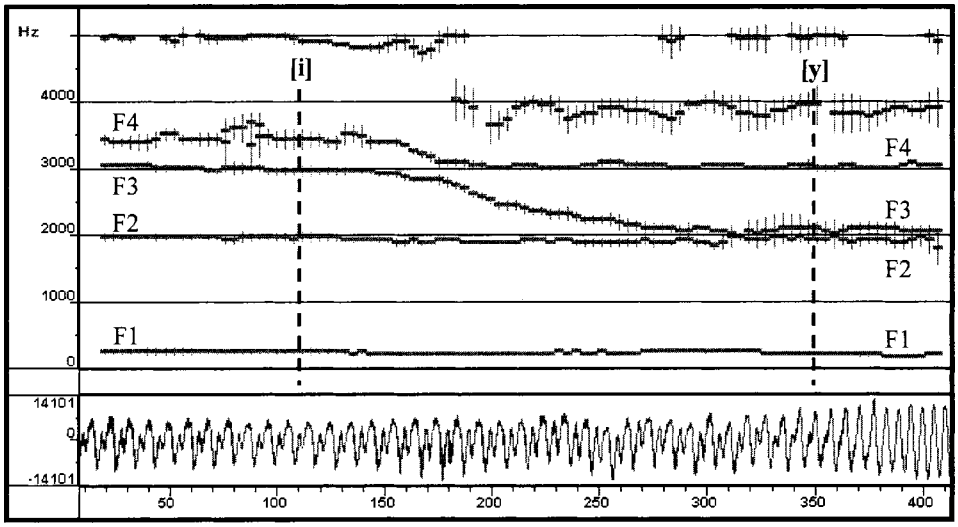


FIG. 6. Formant tracking in the [iy] transition of speaker CS. (Computed with Winsnoori, Babel Technologies, Belgium.)

ing in an [iy] sequence: the first resonance frequency of the front cavity, represented for [y] by $F2$, originates from the fourth formant $F4$ of [i]; consequently, formants $F2$ and $F3$ of [i] are, respectively, the half-wavelength and the wavelength resonance modes of the back cavity, which are in turn represented by $F3$ and $F4$ for [y]. Using a similar approach, it is possible to identify the formant associated with the half-wavelength resonance of the back cavity of [a] by examining the trajectory of the second formant of [i] during an [ia] sequence.

IV. EVALUATION OF THE METHOD

The RBM aims at accounting for interspeaker variability by integrating morphological and articulatory differences between speakers. This method requires us to know which vowel is produced, in order to use the appropriate cavity length ratios. Hence, it is well adapted to speaker transformation purposes, where the objective is to transform a sentence pronounced by a speaker A so that it sounds like a sentence pronounced by speaker B. Interesting studies have been proposed in the last 30 years that contributed to this objective (e.g., Childers *et al.*, 1989; Nordström, 1975, 1977; Story and Titze, 2002; Titze *et al.*, 1996, 1997; Wong *et al.*, 1996). However, for all of them the problem of a quantitative evaluation of their efficiency was never really solved. Therefore, the evaluation framework that we chose for the RBM consisted of transforming five male speakers into the reference articulatory model of the vocal tract, and in measuring the corresponding reduction of the dispersion in the $(F1,F2)$ plane of the formant distribution measured for the whole set of speakers. It permits a quantitative comparison of the RBM with normalization techniques published in the literature.

A. Corpus

Since the reference articulatory model of the vocal tract was built using data from a male speaker without any additional transformation (cf. Sec. III A), we recorded five male speakers. They were all native speakers of French, agreed to participate voluntarily in our experiment, and none of them had any record of pathology of speech production or of the auditory system.

The corpus was designed according to three major requirements for the vowels to be analyzed: (1) reducing the acoustical coupling between the back and front vocal-tract cavities; (2) reducing the token-to-token intraspeaker variability; (3) favoring for each speaker the production of extreme articulations for cardinal vowels [i, a, u]. To this end, the subjects were asked to pronounce the vowels within CVC sequences, where C is a constant, in order to favor the production of the closest possible configuration of vowel V, and thus to reduce the coupling. In addition, consonant C was a voiced consonant, in order to facilitate formant tracking. Each sequence was pronounced within a short word (given in Table III), which was in turn embedded in the carrier sentence: "C'est *CVC* ça?" ("That's *CVC*?"). Subjects were required to repeat each sentence ten times.

In addition to this corpus, the vowel transitions [i–y], [i–a], and [a–i] were also recorded. These transitions were used for each subject to clarify his formant affiliations for the focal vowels [i], [y], and [a] (cf. Sec. III B). They were pronounced in the same carrier sentence as above and each subject repeated them ten times.

The sound recording was carried out in an interactive environment, in an anechoic chamber. The speakers, seating in an armchair, had to read the items displayed on a PC screen. At the beginning of the session, the subjects read the first two phrases of the CVC corpus without being recorded, so that they could get familiar with the environment. Then, for each phrase the acquisition started automatically, as soon as the speech signal level exceeded an experimentally fixed threshold of -24 dB. The speakers were instructed to always favor pronunciation clarity. After the recordings a perceptual verification was performed in order to ensure that the tokens produced by the speakers corresponded well to the desired phonetic category.

The speech signal was captured by a dynamic micro-

TABLE III. CVC[V] items of the closed-context vowel corpus.

Vowel	[i]	[y]	[u]	[e]	[o]	[ɛ]	[ɔ]	[a]
Word	*zizi*	*juju*	*gougou*	*zézé*	*gaugau*	*zêzê*	*troc*	*rara*

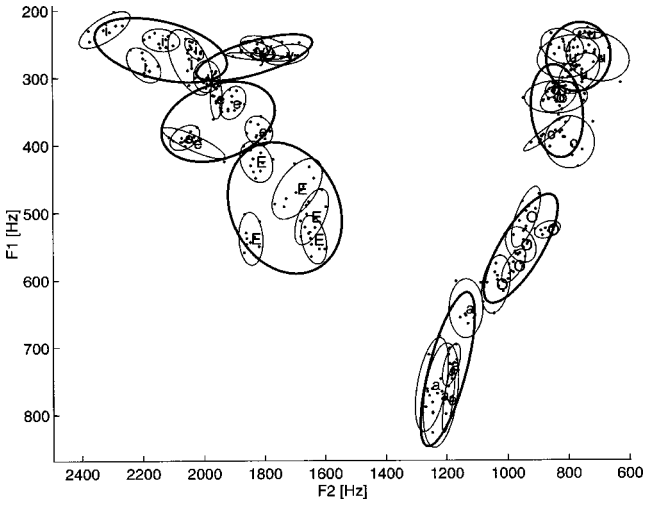


FIG. 7. Vowel distributions in the $(F1, F2)$ plane for the five speakers before transformation using the RBM. The bold line ellipses characterize the dispersions for the whole set of speakers, while individual dispersions are represented with thin line ellipses.

phone [*Beyerdynamic M 10 1N(C)*], amplified by a mixer-amplifier (*Yamaha MV 802*), and sampled at a frequency of 20 000 Hz, using a soundboard (*Oros*) installed on a PC.

Formants were detected with an LPC analysis as follows: 20 coefficients, Hamming window of 20 ms; shift of the window: 5 ms; signal pre-emphasis (coefficient: 0.95). After phoneme labeling, the formants of each vowel were measured in the middle of its steady zone. The formant values used henceforth for the evaluation of the method correspond to the central vowel of the CVC sequences.

Figure 7 shows the vowel distributions for the whole set of speakers in the $(F1,F2)$ plane. The bold line ellipses characterize the dispersions for the whole set of speakers, while individual dispersions are represented with thin line ellipses.

B. Formant affiliations proposed for the real speakers

As explained above, the RBM uses the lowest half-wavelength and quarter-wavelength resonance frequencies or, in some cases, the Helmholtz frequency, in order to infer the ratios of the cavity lengths between two speakers (cf. Sec. II B). The following notations will be used henceforth:

(i) $R1$ is the first resonance mode of the back cavity; usually, it is a Helmholtz resonance of the set ''back cavity+constriction;''
(ii) $R2$ is the lowest resonance of the front cavity, which may be either a quarter-wavelength resonance, if the lips are open, or a Helmholtz resonance, if lips are rounded;
(iii) $R3$ is the second resonance mode of the back cavity, which is in all cases a half-wavelength one.

The formant affiliations proposed for the human speakers are given in Table IV. For [i, a, y] these associations were determined using the vowel transitions (see Sec. III D), whereas for the other vowels the patterns established for the reference vowels (see Sec. III B 2) were taken into consideration. When two affiliations seemed equally possible, we took into account *a posteriori* the affiliation leading to the best results.

TABLE IV. Formant–resonance associations proposed for the vowels of the human speakers. Legend: $R1$ and $R3$ are, respectively, the first and the second resonance modes of the back cavity; $R2$ is the lowest resonance of the front cavity.

Vowel	$R1$	$R2$	$R3$
[a]	$F1$	$F2$	$F4$
[i]	$F1$	$F4$	$F2$
[u]	$F1$	$F2$	$F3$
[o]	$F1$	$F2$	$F3$
[ɔ]	$F1$	$F2$	$F3$
[e]	$F1$	$F3$	$F2$
[y]	$F1$	$F2$	$F3$
[ɛ]	$F1$	$F2$	$F3$

C. Reduction of variability among speakers

The evaluation of the method consisted of two steps. First, the vowel formants of each human speaker were transformed into the formant space of the reference articulatory model of the vocal tract. Then, a comparison was made in the $(F1,F2)$ plane between the scattering of the points before and after transformation. A reduction of this dispersion would attest to the effectiveness of the procedure.

The speaker transformation follows the successive steps of the procedure given in Secs. II B and III C. First, for the back and the front cavity the length ratios between speakers were calculated for the *basis vowels* according to the formulas

$$\alpha_b[v]=R3_M/R3_X, \quad v\in\{\text{a, i, u}\}, \tag{9}$$

$$\alpha_f[v]=R2_M/R2_X, \quad v\in\{\text{a, i}\}, \tag{10}$$

and

$$\alpha_f[u]=\left(\frac{R2_M[u]}{R2_X[u]}\right)^2, \tag{11}$$

where M stands for *reference articulatory Model of the vocal tract* and X for the real speaker to be transformed; $R2$ and $R3$ have the meanings given in Sec. IV B. The $\alpha_f[u]$ expression takes into account the fact that for this vowel, $R2$ is a Helmholtz resonance. The α ratios of the other vowels were determined by interpolation between the respective values of the *basis vowels* (cf. Sec. III C). The ''transformed'' resonances of speaker X were obtained for each vowel as follows:

(i) For a half- or a quarter-wavelength resonance, the transformed value was calculated as the product between the initial value and the α ratio corresponding to the associated cavity.
(ii) For a Helmholtz resonance

$$Ri_N[v]=Ri[v]\sqrt{\alpha_c[v]}, \tag{12}$$

where $(i,c)\in\{(1,b);(2,f)\}$.

The front and back cavity length ratios obtained with this method for the eight vowels and for the five speakers are

TABLE V. Values of the length ratios of the front (F) and back (B) vocal-tract cavities for the eight vowels and the five speakers.

	[i]		[y]		[e]		[ɛ]		[a]		[ɔ]		[o]		[u]	
Speaker	F	B	F	B	F	B	F	B	F	B	F	B	F	B	F	B
CS	0.84	1.02	0.83	1.04	0.84	1.03	0.82	1.04	0.97	0.97	0.92	1.03	0.75	1.07	0.74	1.07
JLS	0.77	0.94	0.78	0.94	0.78	0.94	0.79	0.94	0.95	0.96	0.93	0.94	0.85	0.93	0.85	0.93
MP	0.79	0.97	0.80	1.02	0.79	0.99	0.81	1.03	1	1.	0.98	1.06	0.92	1.1	0.92	1.1
PB	0.75	1.01	0.75	1.02	0.75	1.02	0.76	1.03	1.05	0.98	0.99	1.02	0.79	1.04	0.78	1.04
JMD	0.83	0.89	0.85	0.96	0.84	0.92	0.86	0.97	0.99	1.03	1.	1.06	1.	1.08	1.	1.08

given in Table V. It can be seen that noticeable differences exist for each speaker between the length ratios of the back and front cavities. This supports our strategy of using two length ratios; a unique length factor applied to the whole vocal-tract length would not be able to account for the specific vocal-tract geometry of each speaker. A principal component analysis applied to these data reinforces this statement: two factors are at least required to describe more than 80% of the length ratios variance (the first three factors, respectively, explain 57.6%, 30.6%, and 10.6% of the data variance). In addition, clear differences can be observed between the length ratios obtained for [i], [a], and [u]. This supports our decision to use a specific ratio for each extreme cardinal vowel.

The reduction of variability obtained with the length ratios of Table V was quantitatively measured by calculating for each vowel in the $(F1, F2)$ plane the area of the 2σ dispersion ellipsis, before and after transformation. The results are given in Table VI. Note that the areas are not given in Hz^2, but in percentage. This percentage was calculated as the ratio between the ellipsis area and the area of the rectangle defined by the $F1$ and $F2$ ranges of the global vowel distribution of the five speakers. This was necessary for further comparison with normalization techniques. The dispersion of the vowels obtained for the whole set of speakers after the formants transformation provided by the RBM method is given in Fig. 8. As in Fig. 7, the bold line ellipses characterize the dispersions for the whole set of speakers, while individual dispersions are represented with thin line ellipses.

One can note that except for vowels [o] and [y], the global variability clearly decreases after speaker transformation using the RBM. The clear general reduction of the variability suggests that, in spite of its underlying simplifying hypotheses, the RBM accounts fairly well for the main causes of the interspeaker variability. At the same time, for [y] no noticeable reduction is observed, and for [o] the variability increases clearly. In both cases, the first two formants are Helmholtz resonances. The lack of reduction of the variability could then arise from the fact that focusing on tube lengths variability could be inaccurate, in the cases where, as in Helmholtz resonators, the resonance frequency depends on cavity volumes.

TABLE VI. Areas of the 2σ dispersion ellipses, before and after speaker transformation provided by the RBM. The values are given in percentage with respect to the area of the smallest rectangle including the whole vowel space in the $(F1,F2)$ plane.

	[a]	[i]	[u]	[o]	[ɔ]	[e]	[ɛ]	[y]
Initially	20.90	27.18	15.52	16.85	18.24	30.62	51.90	14.08
RBM	16.77	9.61	8.50	22.06	16.00	12.51	44.95	14.03

Considering the remaining six vowels, it can be observed that the variability decrease after transformation is larger for closed than for open vowels. According to the theoretical hypotheses underlying the RBM, this trend can be explained by the fact that for open vowels interspeaker variability in constriction size and, then, in acoustical coupling between cavities has a non-negligible influence on formant variability.

D. Comparison with speaker normalization techniques

The transformation of the formant patterns of five speakers into a reference formant space can be compared to a classical normalization procedure. Thus, it offers a good framework to evaluate the relevance of the RBM, by comparing the obtained reduction of speaker variability with the ones generated by six normalization techniques published in the literature. Again, it should be emphasized that the RBM cannot be used in a pure normalization framework, since this method requires one to know which vowel is pronounced before it can be transformed into the formant space of another speaker.

Among the considered normalization techniques, four take into account statistical properties of the dispersion of each vowel across speakers. These were proposed by Gerstman (1968), Lobanov (1971), Nearey (1977) (see also

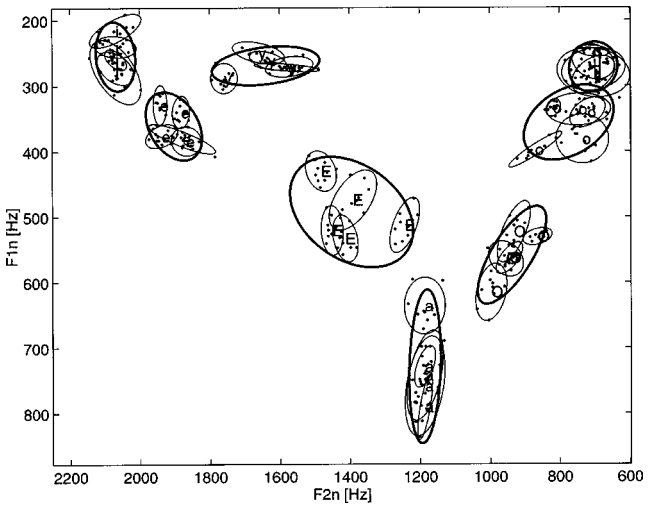


FIG. 8. Vowel distributions in the $(F1, F2)$ plane for the five speakers after transformation using the RBM. See Fig. 7 for legend.

TABLE VII. Areas of the 2σ dispersion ellipses for all the normalization procedures, before and after normalizations. The values are given in percentage with respect to the area of the smallest rectangle including the whole vowel space in the $(F1,F2)$ plane.

	[a]	[i]	[u]	[ɒ]	[ɔ]	[e]	[ɛ]	[y]
Initially	20.90	27.18	15.52	16.85	18.24	30.62	51.90	14.08
Gerstman	18.59	9.38	12.76	23.86	25.82	15.79	74.23	13.15
Lobanov	21.99	11.35	12.88	14.91	21.96	11.95	45.20	14.35
Nearey	12.24	25.33	42.40	29.71	17.06	20.18	34.42	16.64
Di Benedetto	0.00	0.00	0.00	21.14	26.00	20.96	75.31	19.14
Miller (Hz)	21.51	27.68	37.95	43.94	24.89	12.20	30.30	7.20
Miller (Mel)	18.74	22.96	42.14	47.39	23.27	10.81	25.88	7.11
RBM	16.77	9.61	8.50	22.06	16.00	12.51	44.95	14.03

Nearey, 1980), and Di Benedetto and Liénard (1992). The other two techniques were elaborated by Miller (1989). They use nonlinear scale frequencies based on perceptual criteria.

In its original form, Gerstman's (1968) procedure consisted of a homothetic transform that maps the initial variation ranges of $F1$ and $F2$ into a normalized domain, [0, 999] Hz. We implemented a strictly equivalent version of this method (Ferrari Disner, 1980), in which the limits of the normalized $(F1, F2)$ plane are, more realistically, [250, 750] Hz for $F1$, and [850, 2250] Hz for $F2$.

Lobanov's method (1971) normalizes the formant values of each speaker according to the relation

$$F_i^N=(F_i-M_i)/\sigma_i, \quad i\in\{1,2\}, \tag{13}$$

where F_i^N is the normalized value of the F_i formant, and M_i and σ_i are, respectively, the mean value and the standard deviation of F_i calculated for the whole set of vowels.

Nearey (1977) suggested using a logarithmic transformation

$$F_{ijk}^N=\log(F_{ijk})-F_{\text{mean},k}, \tag{14}$$

where i, j, and k represent the formant, the vowel, and the speaker, respectively. $F_{\text{mean},k}$ is the mean of the logarithmic transforms of the first two formants calculated for all the vowels of speaker k.

Di Benedetto and Liénard (1992) proposed to project, through a linear transformation, the $(F1, F2)$ plane of a given speaker, onto a ''standard'' vowel plane defined by the mean values of $F1$ and $F2$ computed for the extreme vowels [a, i, u] produced by a large number of speakers. In concrete terms, for a given speaker, the normalized values of the first two formants of a vowel v, are calculated as follows.

(1) The position of vowel v in the speaker's $(F1, F2)$ plane is expressed as a linear function of the positions of his three extreme vowels [i], [a], and [u] with an *ad hoc* constraint applied to the sum of the weights used in the linear formula. This is expressed by the matrix product

$$\begin{pmatrix} F1[a] & F1[i] & F1[u] \\ F2[a] & F2[i] & F2[u] \\ 1 & 1 & 1 \end{pmatrix}\cdot\begin{pmatrix} \alpha[u] \\ \beta[v] \\ \gamma[v] \end{pmatrix}=\begin{pmatrix} F1[v] \\ F2[v] \\ 1 \end{pmatrix}, \tag{15}$$

summarized by the relation $A\cdot\theta[v]=V$ (16)

(2) Each vowel v in the speaker's $(F1, F2)$ plane is transformed into a normalized vowel v_0 with a linear function. Consequently, the relative position of v_0 in the normalized domain relatively to the extreme vowels [a], [i], and [u] is the same as the position of v in the speaker's domain, and it is possible to write

$$V_0=A_0\cdot\theta[v], \tag{17}$$

where A_0, structured like A, contains the mean values of the first two formants of the vowels [a, i, u] calculated over several speakers.

(3) Since $\theta[v]$ is the same in the two spaces, the normalized $F1$ and $F2$ values for the vowel $[v]$ can be deduced as follows:

$$V_0=A_0\cdot A^{-1}\cdot V. \tag{18}$$

Miller (1989) evaluated the capability of several nonlinear frequency scales to reduce the general interspeaker variability, in particular between groups of men–women–children. The two most efficient transformations were selected for our evaluation, namely $\log_{10}(F2/F1)$ and $\log_{10}(F3/F2)$ expressed first in Hz, and then in Mel.

Table VII allows the comparison of the different techniques described above. The input data were the vowels produced by the real speakers. As in Sec. IV C, the areas of the 2σ dispersion ellipsis, normalized with respect to the size of the considered vowel space, are given. The first line of Table VII gives the initial dispersions of each sound class before normalization. Figures 9–14 illustrate the changes of the

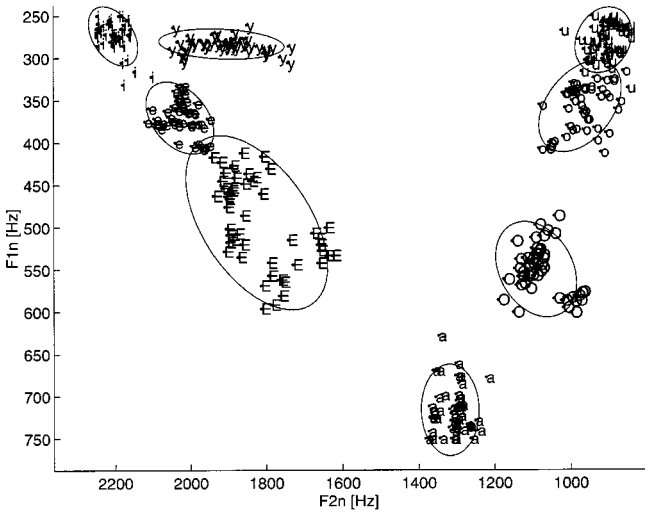


FIG. 9. Vowel distributions in the $(F1. F2)$ plane for the five speakers after normalization with Gerstman's method.

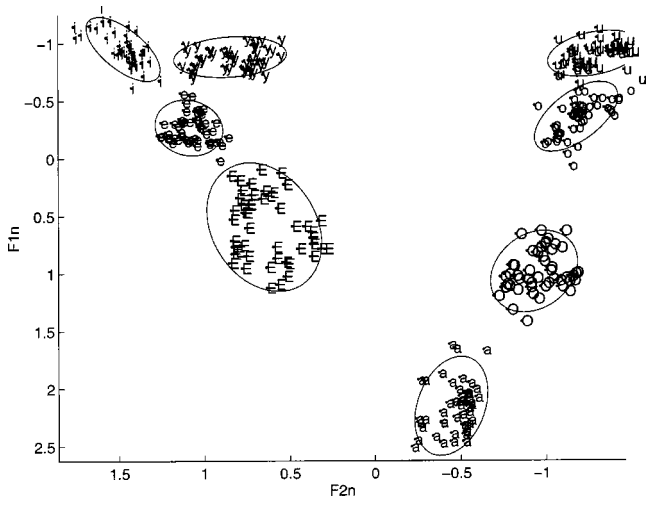


FIG. 10. Vowel distributions in the $(F1, F2)$ plane for the five speakers after normalization with Lobanov's method.

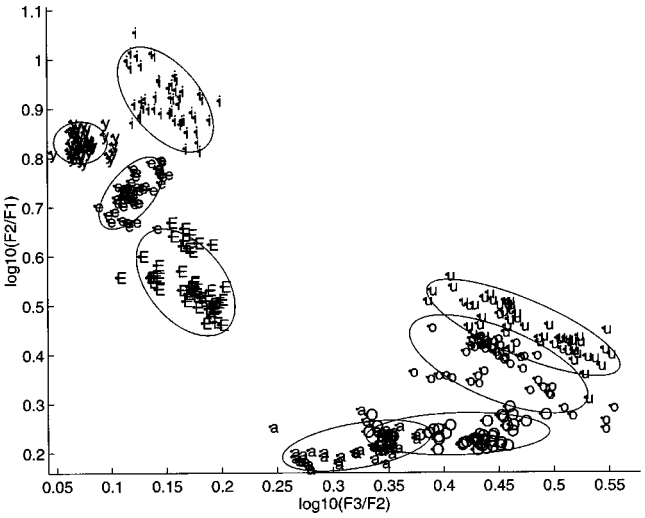


FIG. 13. Vowel distributions in the $(F1, F2)$ plane for the five speakers after normalization with Miller's method in Hz.

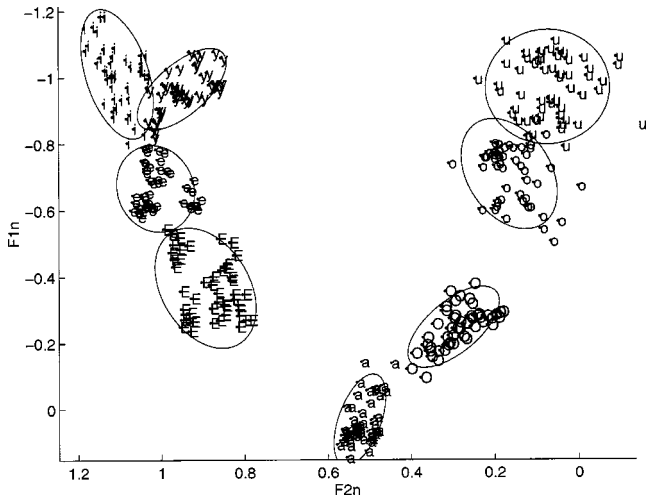


FIG. 11. Vowel distributions in the $(F1, F2)$ plane for the five speakers after normalization with Nearey's method.

acoustic space after application of each of the normalization methods. They can be compared with Fig. 8, which shows the results of the transformation with the RBM. The RBM is, thus, not systematically the most successful procedure, but in almost all cases it is one of the most efficient methods. A more quantitative assessment was made by establishing a ranking of the different procedures computed on the basis of their average ellipsis area. It should be noted that the rankings established for [a], [i], and [u] do not take into account Benedetto and Liénard's procedure, since it intrinsically eliminates the variability for these extreme vowels. The results are given in Table VIII. It can be seen that the RBM is the method that generates on average the largest variance reduction. This result is especially positive, since, contrary to the normalization methods, the transformation provided by the RBM is based on information taken from only three vowels, the *basis* vowels, and considering only the first three formants.

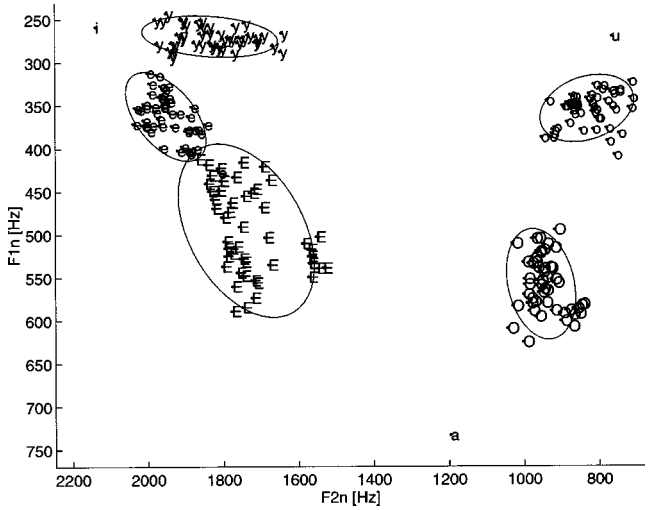


FIG. 12. Vowel distributions in the $(F1, F2)$ plane for the five speakers after normalization with Di Benedetto and Liénard's method.

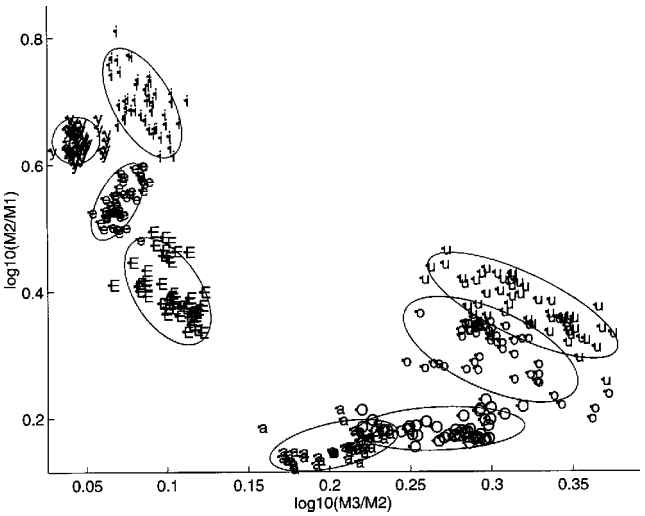


FIG. 14. Vowel distributions in the $(F1, F2)$ plane for the five speakers after normalization with Miller's method in Mel.

TABLE VIII. Average ellipsis areas computed for all vowels (for Di Benedetto and Lienard's method see the text).

RBM	Lobanov	Gerstman	Nearey	Miller (Mel)	Miller (Hz)	Di Benedetto
18.054	19.324	24.198	24.748	24.788	25.709	32.51

V. DISCUSSION

The first conclusion of our study is that interpreting the relations between vocal-tract geometry and formant frequencies by using the formant–cavity affiliation concept is an efficient approach to give a good account of interspeaker variability. These results support our hypothesis that considering separately the length of the back and of the front vocal-tract cavities is an interesting method to model the origins of interspeaker variability. However, beyond this first positive statement, the detailed comparison of Figs. 7 and 8 shows that some aspects of the results obtained with the RBM are not fully satisfactory.

A. Limitations of the speaker transformation based on the RBM

Comparing Figs. 7 and 8, it appears that for vowels [i], [e], [ɛ], [ɔ], and [a] the decrease of the dispersion in the $(F1, F2)$ plane is mainly due to a reduction of the $F2$ variability range, while no noticeable change is observed along the $F1$ axis. For these five vowels, the clear decrease of the $F2$ variability suggests that the affiliations and the corresponding values of the length ratios proposed in the RBM are correct. Now, how can we explain the very weak impact of our speaker transformation procedure on $F1$? Two explanations can be proposed. The RBM assumes that $F1$ is a Helmholtz resonance of the set "back cavity + constriction." Therefore, a first possible explanation would be that, contrary to our assumption, changes in the cross-sectional plane would contribute significantly to the variation of $F1$. In such a case, modeling changes in cavity length would only partially account for the interspeaker variability of $F1$. An alternative explanation could be that, at least for vowel [a], $F1$ would not be a Helmholtz resonance (see below for further analysis).

For vowel [u], Fig. 8 shows a clear decrease of the variability both in $F1$ and $F2$ directions, but this is not the case for the other rounded vowels [y] and [o], especially along the $F1$ axis. Here again, two possible explanations can be proposed. As said above, since for these two vowels $F1$ and $F2$ are considered to be Helmholtz resonances, taking into account changes in the cross-sectional plane in order to integrate volume changes could be a necessity in order to model the interspeaker variability with enough accuracy. An alternative or additional explanation could be that the association between formants and resonances was not correctly established for these two vowels (see below for further analysis).

Thus, in addition to the already mentioned possible non-negligible coupling between cavities, it appears that errors in the association between formants and resonances could contribute to the limitation of the efficiency of the RBM. These errors can be due to an inadequate acoustic model for vowel production or to difficulties in establishing the correct affiliations in the framework of Fant's four-tube model.

B. An alternative acoustic model for some vowels?

Fant's four-tube model and its associated vocal-tract nomograms, on which we based on RBM, do constitute a generally well-accepted reference for the understanding of vowel acoustics. Nevertheless, alternative well-known models have also been proposed in the literature, which would lead to somewhat different hypotheses. In particular, Stevens (1972, 1989, 1998) suggested modeling the vowel area functions only with three tubes and even with two tubes, when the vowel is articulated either in the very back or the very front region of the vocal tract. Thus, for vowel [a], Stevens' model consists of the concatenation of a narrow tube accounting for the constriction, and of a larger tube representing the front cavity (see Stevens, 1998, p. 274). For very front vowels like [i], Stevens (1989, 1998) suggested using a three-tube model (see Stevens, 1998, p. 277). Consequently, both for [a] and for [i], the affiliation patterns suggested by Stevens are different from those that were taken into consideration in the RBM on the basis of the four-tube model. The affiliations suggested by Stevens for the first four formants of [i] and [a] are given in Table IX.

Thus, for [a], the lowest resonance affiliated with the back cavity would be a quarter-wavelength resonance instead of a Helmholtz resonance (see also Mrayati and Carré, 1976; Bailly, 1993). Considering this suggestion for some speakers in the RBM could possibly enhance the decrease of the $F1$ range after transformation for [a].

For vowel [i], Stevens' three-tube modeling is associated with a shortening of the front cavity and a lengthening of the constriction (Stevens, 1998, p. 277). In this case, the third formant would not be a quarter-wavelength resonance of the front cavity, but rather a half-wavelength resonance of the constriction tube. Here again, it would be interesting to observe the consequences of this suggestion on the RBM.

Both kinds of models are generally considered to be plausible.[4]

C. Inaccuracy of the models in focal regions

Another limitation of the RBM arises from the fact that the extreme vowels that serve as *basis vowels* in our method

TABLE IX. Association between formants and vocal tract resonances for [a] and [i] as suggested by Stevens' model (1989, 1998), Legend: *H. back* = Helmholtz resonance of the set "back cavity + constriction;" for the other resonances: *back* = back cavity; *front* = front cavity.

Vowel	$F1$	$F2$	$F3$	$F4$
[a]	λ/4 front	λ/4 back	3λ/4 front	3λ/4 back
[i]	H. back	λ/2 back	λ/2 constriction	λ back

are articulated in focal regions of the vocal tract. Such regions appear both on the nomograms generated with the four-tube model (Fant, 1960; Badin *et al.*, 1990), and on those obtained with the two- and/or three-tube model (Stevens, 1972, 1989, 1998). In the case of a two-tube model, focal points correspond to the merging of the quarter-wavelength resonances of the front and of the back cavities. In the case of three and four tubes, if the constriction tube is long enough, a focal region can correspond to the merging of up to three resonances. Therefore, for a configuration articulated in such a region, there is an uncertainty concerning the affiliations between formants and cavities, even if the coupling between cavities is very small. When the coupling increases, the uncertainty zone widens on both sides of the focal point.[5] For a vowel articulated in such a region, affiliations can be properly determined only if one knows on what side of the focal point the vowel is.

Given the interpolation functions that are used in the RBM, errors in formants–cavities affiliation patterns for *basis* vowels could generate errors in the whole $(F1, F2)$ plane. Therefore, looking for possibilities to increase the accuracy of the determination of these patterns would noticeably increase the reliability of the RBM.

VI. CONCLUSION

Our original hypothesis was that interspeaker variability of the formant patterns for oral vowels originates in differences in the lengths of the back and front vocal-tract cavities, and that it is possible to infer them from an analysis of the formants in terms of vocal-tract cavity resonances. Our results support this assumption, since the *resonance-based method* permits us to give a fair account of the specificity of each speaker's speech production. In particular, the generally good reduction of the variability along the $F2$ axis supports the validity of the concept of affiliation between formants and cavities, as well as of the majority of the affiliations that were chosen in the procedure. It also supports the hypothesis that the influence of interspeaker differences in acoustical coupling between cavities can be neglected as compared to the influence of differences in cavity lengths.

However, limitations of the method were also shown, which could be linked either to an inexact determination of some affiliations, especially in the focal regions, or to the fact that interspeaker geometrical differences in the cross-sectional plane cannot be neglected. Considering an alternative model of vowel production, such as Stevens' (1998) model, could contribute to solve at least a part of the inappropriate hypotheses about the affiliations. Collecting 3D geometrical data on a number of speakers and studying the associated acoustic signal would permit us to evaluate quantitatively the influence on the acoustics of the variability in the cross-sectional plane.

In general, the RBM was shown to be an interesting tool to understand and model interspeaker variability in vowel production, when the subjects are recorded in laboratory conditions and in a consonantal context that favors the most prototypical articulation of the vowels. Working on connected speech, which is usually hypoarticulated, and for which more acoustical coupling between cavities is to be expected, should permit us to assess the practical usefulness of the RBM for speech technology.

ACKNOWLEDGMENTS

This work was supported by doctoral scholarships from the *Robert Schuman Foundation* and from the *Région Rhônes-Alpes (Program TEMPRA)* to the first author.

[1]The idea that a formant may be considered as a characteristic resonance of a vocal-tract cavity had already been suggested (cf. Chiba and Kajiyama, 1941; Dunn, 1950).

[2]This model takes into account all the boundary conditions: the heat conduction and the viscosity losses are considered with unitary shape factor; the radiance at the lips is modeled by a piston in an infinite plane, and the wall vibrations are accounted for by localized impedances along the vocal tract, independently of the area function. There is no subglottal coupling.

[3]This inversion is based on a gradient technique that has systematically as an initial condition the neutral shape of the vocal tract (for more details, cf. Bailly *et al.*, 1995).

[4]Note that Fant himself sometimes used a two-tube modeling, in particular for [i] (e.g., Fant, 1966, 1975).

[5]Indeed, the more the coupling is important, the more the slope of the formants' trajectories moves away from that of the resonances of the decoupled cavities, and comes close to zero (cf., e.g., Fant, 1960, p. 77).

Abry, C., and Boë, L.-J. (**1986**). "Laws for lips," Speech Commun. **5**, 97–104.

Badin, P., and Fant, G. (**1984**). "Notes on vocal tract computations," *Speech Transmission Laboratory—Quarterly Progress and Status Report* (Royal Institute of Technology, Stockholm), **2–3**, pp. 53–108.

Badin, P., Perrier, P., Boë, L.-J., and Abry, C. (**1990**). "Vocalic nomograms: Acoustic and articulatory considerations upon formant convergence," J. Acoust. Soc. Am. **87**(3), 1290–1300.

Bailly, G. (**1993**). "Resonances as possible representation of speech in the auditory-to-articulatory transform," in Proceedings of the 3rd European Conference on Speech Communication and Technology (Eurospeech 93), Berlin, Germany, Vol. 2, pp. 1511–1514.

Bailly, G., Boë, L.-J., Vallée, N., and Badin, P. (**1995**). "Articulatory-acoustic vowel prototypes for speech production," in Proceedings of the 4th European Conference on Speech Communication and Technology (Eurospeech 95), Madrid, Spain, Vol. 3, pp. 1913–1916.

Beautemps, D., Badin, P., Bailly, G., Galván, A., and Laboissière, R. (**1996**). "Evaluation of an articulatory-acoustic model based on a reference subject," 1st ESCA Tutorial and Research Workshop on Speech Production Modelling—4th Speech Production Seminar, pp. 45–48.

Boë, L.-J., Perrier, P., and Bailly, G. (**1992**). "The geometric vocal tract variables controlled for vowel production: Proposals for constraining acoustic-to-articulatory inversion," J. Phonetics **20**, 27–38.

Bothorel, A., Simon, P., Wioland, F., and Zerling, J.-P. (**1986**). *Cinéradiographie des Voyelles et des Consonnes du Français (Cineradiography of French Vowels and Consonants)* (Travaux de l'Institut de Phonétique de Strasbourg: Université des Sciences Humaines, Strasbourg, France).

Chiba, T., and Kajiyama, M. (**1941**) (reprinted in 1958). *The Vowel—Its Nature and Structure* (Tokyo, Japan).

Childers, D. G., Wu, K., Hicks, D. M., and Yegnanarayana, B. (**1989**). "Voice conversion," Speech Commun. **8**(2), 147–158.

Di Benedetto, M. G., and Liénard, J.-S. (**1992**). "Extrinsic normalization of vowel formant values based on cardinal vowels mapping," in Proceedings of the 2nd International Conference on Spoken Language Processing, Banff–Alberta, Canada, pp. 579–582.

Dunn, H. K. (**1950**). "The calculation of vowel resonances, and an electrical vocal tract," J. Acoust. Soc. Am. **22**, 740–753. [Reprinted in: Flanagan, J. L., and Rabiner, L. R., editors (**1973**). *Speech Synthesis* (Dowden, Hutchinson, and Ross, Stroudsburg, PA.]

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague, The Netherlands).

Fant, G. (**1966**). "A note on vocal tract size factors and non-uniform F-pattern scalings," *Speech Transmission Laboratory—Quarterly Progress and Status Report* (Royal Institute of Technology, Stockholm), **4**, pp. 22–30.

Fant, G. (**1975**). ''Non-uniform vowel normalization,'' *Speech Transmission Laboratory—Quarterly Progress and Status Report* (Royal Institute of Technology, Stockholm), **2–3**, pp. 1–19.
Fant, G., and Pauli, S. (**1974**). ''Spatial characteristics of vocal tract resonance modes,'' Speech Commun. Seminar **2**, 121–132.
Ferrari Disner, S. (**1980**). ''Evaluation of vowel normalization procedures,'' J. Acoust. Soc. Am. **67**(1), 253–261.
Gerstman, L. J. (**1968**). ''Classification of self-normalized vowels,'' IEEE Trans. Audio Electroacoust. **AU-16**, No. 1, 78–80.
Klatt, D. (**1986**). ''The problem of variability in speech recognition and in models of speech perception,'' in *Invariance and Variability in Speech Processes*, edited by J. S. Perkell and D. H. Klatt (Erlbaum, Hillsdale, NJ), pp. 300–319.
Lobanov, B. M. (**1971**). ''Classification of Russian vowels spoken by different speakers,'' J. Acoust. Soc. Am. **49**(2), 606–608.
Maeda, S. (**1990**). ''Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model,'' in *Speech Production and Speech Modelling*, edited by W. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 131–149.
Majid, R., Abry, C., Boë, L.-J., and Perrier, P. (**1987**). ''Contribution à la classification articulatori-acoustique des voyelles: Étude des macrosensibilités à l'aide d'un modèle articulatoire'' (''A contribution to vowel classification in the articulatory and acoustic domains: Studying macrosensitivities with an articulatory model''), in *Proceedings of the 11th International Congress of Phonetic Sciences* (Tallin, Estonia), Vol. 2, pp. 348–351.
Miller, J. D. (**1989**). ''Auditory-perceptual interpretation of the vowel,'' J. Acoust. Soc. Am. **85**(5), 2114–2134.
Mrayati, M., and Carré, R. (**1976**). ''Relations entre la forme du conduit vocal et les caractéristiques acoustiques des voyelles françaises—Etude des distributions spatiales'' (''Relations between vocal tract geometry and acoustic characteristics for French vowels—A study of spatial distributions''), Phonetica **33**, 285–306.
Nearey, T. (**1977**). ''Phonetic feature systems for vowels,'' Doctoral dissertation, University of Connecticut, Storrs, CT. Reproduced by Indiana University Linguistics Club, 1978. Unavailable. Cited by Ferrari Disner (1980).
Nearey, T. (**1980**). ''On the physical interpretation of vowel quality: Cinefluorographic and acoustic evidence,'' J. Phonetics **8**, 213–241.
Nordström, P. E. (**1975**). ''Attempts to simulate female and infant vocal tracts from male area functions,'' *Speech Transmission Laboratory—Quarterly Progress and Status Report* (Royal Institute of Technology, Stockholm), **2–3**, pp. 20–33.
Nordström, P. E. (**1977**). ''Female and infant vocal tracts simulated from male area functions,'' J. Phonetics **5**, 81–92.
Nordström, P. E., and Lindblom, B. (**1975**). ''A normalization procedure for vowel formant data,'' in Proceedings of the 8th International Congress of Phonetic Sciences in Leeds, Paper 212.
Perkell, J. S., and Klatt, D. H., (editors) (**1986**). *Invariance and Variability in Speech Process* (Erlbaum, Hillsdale, NJ).
Savariaux, C., Perrier, P., and Orliaguet, J. P. (**1995**). ''Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production,'' J. Acoust. Soc. Am. **98**(5), Pt. 1, 2428–2442.
Stevens, K. N. (**1972**). ''The quantal nature of speech: Evidence from articulatory-acoustic data,'' in *Human Communication: A Unified View*, edited by E. E. David, Jr. and P. B. Denes, McGraw-Hill, New York), pp. 51–66.
Stevens, K. N. (**1980**). ''Acoustic correlates of some phonetic categories,'' J. Acoust. Soc. Am. **68**(3), 836–842.
Stevens, K. N. (**1989**). ''On the quantal nature of speech,'' J. Phonetics **17**, 3–45.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (The MIT Press, Cambridge, MA).
Story, B. H., and Titze, I. R. (**2002**). ''A preliminary study of voice quality transformation based on modifications to the neutral vocal tract area function,'' J. Phonetics **30**(3), 485–509.
Straka, G. (**1965**). *Album Phonétique* (*Phonetic Album*) (Presses de l'Université de Laval, Québec, Canada).
Titze, I., Wong, D., Story, B., and Long, R. (**1996**). ''Voice transformation with physiologic scaling principles,'' NCVS Status and Progress Report, **10**, pp. 103–110.
Titze, I., Wong, D., Story, B., and Long, R. (**1997**). ''Considerations in voice transformation with physiologic scaling principles,'' Speech Commun. **22**, 113–123.
Wakita, H. (**1977**). ''Normalization of vowels by vocal-tract length and its application to vowel identification,'' IEEE Trans. Acoust., Speech, Signal Process. **25**(2), 183–192.
Wong, D., Lange, R. C., Long, R. K., Story, B. H., and Titze, I. R. (**1996**). ''Age and gender related speech transformations using linear predictive coding,'' NCVS Status and Progress Report, **10**, pp. 111–126.

Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences

Bronwen G. Evans[a] and Paul Iverson
Department of Phonetics and Linguistics, University College London, 4 Stephenson Way, London NW1 2HE, United Kingdom

(Received 26 December 2002; revised 17 October 2003; accepted 30 October 2003)

Two experiments investigated whether listeners change their vowel categorization decisions to adjust to different accents of British English. Listeners from different regions of England gave goodness ratings on synthesized vowels embedded in natural carrier sentences that were spoken with either a northern or southern English accent. A computer minimization algorithm adjusted F1, F2, F3, and duration on successive trials according to listeners' goodness ratings, until the best exemplar of each vowel was found. The results demonstrated that most listeners adjusted their vowel categorization decisions based on the accent of the carrier sentence. The patterns of perceptual normalization were affected by individual differences in language background (e.g., whether the individuals grew up in the north or south of England), and were linked to the changes in production that speakers typically make due to sociolinguistic factors when living in multidialectal environments. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1635413]

PACS numbers: 43.71.Es, 43.71.Bp [PFA] Pages: 352–361

I. INTRODUCTION

In multidialectal environments (e.g., large cities such as London), native speakers of different accents regularly interact with one another. Speakers in these environments often avoid variants that are markedly regional or unusual, in order to facilitate communication (Trudgill, 1986) and to appear cosmopolitan (Foulkes and Docherty, 1999). However, they also retain some regional variants that show their allegiance to particular social or geographical groups (e.g., Foulkes and Docherty, 1999; Trudgill, 1986; Watt, 1998). In order to understand speech, listeners must somehow tolerate or adjust to this phonetic variation.

In British English, the focus of the present study, vowels are particularly important for distinguishing accents (e.g., Wells, 1982). Vowels can be used, for example, to broadly classify the many regional accents in England as northern or southern (Trudgill, 1986; Upton and Widdowson, 1996). Speakers of southern English accents use the vowel [ʌ] in words such as *buck*, but northern English speakers do not have this vowel; they say *buck* with a higher vowel, [ʊ], such that it becomes a homophone or near-homophone of *book*. Southerners and northerners both use the vowels [a] and [ɑ:], but with somewhat different lexical distributions; words such as *bath*, *dance*, and *ask* are produced with [ɑ:] by southerners and [a] by northerners, even though most words that have these vowels (e.g., *bad* using [a], and *bard* using [ɑ:]) do not differ between accents. When listening to a southerner, native speakers of northern English are thus required to map words that contain [ɑ:] and [ʌ] onto their own lexical representations that may be based on [a] and [ʊ].

Previous research on vowel normalization[1] has primarily examined how listeners adjust to the acoustic consequences of anatomical variation, such as the length of the talker's vocal tract and characteristics of their glottis (e.g., Hillenbrand *et al.*, 1995; Ladefoged and Broadbent, 1957; Nearey, 1989). Listeners have been thought to normalize for anatomical variation by relying on gross acoustic—perhaps language-universal—characteristics of the speech signal such as the fundamental frequency of the vowel and the range of formant frequencies used in a carrier sentence [see Nearey (1989) for a review]. However, accent normalization is different because the realizations of phonemes in an accent cannot be predicted from such basic acoustic factors. That is, accent normalization must require more language-, accent-, and phoneme-specific processes.

How might listeners accomplish accent normalization? Exemplar models of speech perception (e.g., Goldinger, 1996, 1998) have theorized that listeners store phonetically detailed memory traces every time they listen to speech. Johnson (1997) has suggested that these exemplar representations can produce talker normalization effects, if listeners compare the words that they hear to stored exemplars of speech produced by similar talkers. Accent normalization can be viewed as an extreme example of this kind of talker normalization (see also Nygaard and Pisoni, 1998), in that the incoming speech could be compared to stored exemplars of speech produced by talkers with similar accents. For example, individuals who have experience with different British accents may be able to recognize northern-accented speech by mapping it onto similar stored exemplars produced by northern talkers, and recognize southern-accented speech by mapping it onto similar stored exemplars produced by southern talkers. Listeners may thus be able to fully adjust to vowel differences between accents, provided that they have had previous experience with similarly accented speech.

The evidence from cross-language speech research,

[a]Electronic mail: bron@phon.ucl.ac.uk

however, suggests that individuals cannot easily adjust their phonemic categorizations to match the talker, at least when listening to foreign or foreign-accented speech. Novice listeners tend to assimilate foreign phonemes into the same categories that they use for native speech (Flege, 1992; Best, 1994; Best *et al.*, 1988, 2001). Although more experience with a foreign accent improves recognition abilities (e.g., Clarke, 2002), the category assimilation processes are difficult to modify; experienced listeners continue to assimilate most foreign phonemes into native categories, and create new categories for foreign speech primarily in cases where the foreign phonemes are too different from the native phonemes to be strongly assimilated (Flege, 1992, 1995).

Even bilingual listeners do not appear to strongly adjust their phonetic categorization processes when switching between different languages. One could expect, for example, that a Spanish–English bilingual would have a category boundary for /d/-/t/ at a shorter VOT for Spanish than for English, because Spanish speakers produce /t/ with a shorter VOT. However, late or weak bilingual listeners appear to use a single VOT boundary in both languages; they set their VOT boundary to a compromise location between English and Spanish (Flege, 1991, 1992). There is some evidence that early or strong bilinguals adjust their VOT boundaries for different languages (Elman *et al.*, 1977; Flege and Eefting 1987; Hazan and Boulakia, 1992), but the magnitudes of these boundary shifts are small and it is possible that these shifts may be caused by postperceptual processes (Bohn and Flege, 1993). The perception of different accents within the same language could operate similarly to these cross-language cases; listeners may assimilate the incoming speech to the phonetic categories of their own native accent without making specific adjustments for the accent of the speaker (see Flege, 1992) and create new categories only in cases where the non-native phonemes are too different from native phonemes to be strongly assimilated.

The present study investigated whether listeners adjust their vowel categorizations when listening to speech produced with different accents within the same language. The study contrasted two varieties of British English: Sheffield English, a northern variety, and Standard Southern British English (SSBE). Listeners with varying backgrounds were tested: northerners and southerners living in London (Experiment 1) and northerners living in the north of England (Experiment 2). Listeners heard synthesized vowels embedded in natural carrier sentences that were produced in either a Sheffield or SSBE accent. They gave goodness ratings on the vowels, and a computer program iteratively adjusted the F1, F2, F3, and duration values on successive trials until a best exemplar of each vowel was found. The aim was to assess whether listeners change their best exemplar locations based on the accent of the carrier sentence. Of particular interest were the vowels in *bath*, *bud*, and *cud*, which are produced very differently in northern and southern English accents (Wells, 1982).

II. EXPERIMENT 1

Experiment 1 investigated whether listeners from the north and south of England who were living in London adjusted their vowel categorization decisions when listening to speech produced in SSBE and Sheffield English accents. London is a multidialectal community, and anyone living in the city for an extended period of time will have had experience with listening to and interacting with speakers of northern and southern English accents, as well as a wide variety of other accents. All listeners in experiment 1 thus had experience with different English accents, but their own native accents differed in terms of being northern or southern.

A. Method

1. Participants

Twenty-three subjects were tested. All were native English speakers resident in London at the time of testing. They had lived in London for an average of 8.6 years, with a minimum of 1 year. The subjects were 20–45 years old, had no known hearing problems, and reported no speech or language difficulties. Three subjects were dropped from the experiment because their best exemplar locations were not reliable (i.e., their best exemplar locations for vowels that are produced the same in SSBE and Sheffield accents, such as in the words *bird* and *bed*, differed by more than 2 ERB in the two carrier sentences). Of the remaining 20 subjects, 10 were from southern England and 10 were from northern England. This classification of background was based on where they had lived between the ages of 5 and 18 years, an important period for the development of accent (Foulkes and Docherty, 1999).

2. Stimuli and apparatus

The stimuli consisted of synthesized vowels in the phonetic environments /b/-V-/d/, /b/-V-/θ/, and /k/-V-/d/, embedded in naturally spoken recordings of the carrier sentence *I'm asking you to say the word [ ] please*. The /b/-V-/θ/ words were included because northerners and southerners produce *bath* with different vowels. The /k/-V-/d/ words were included in case the potential shift in the *bud* and *cud* vowels with accent was affected by lexical influences (i.e., if these words were produced with the northern [ʊ] vowel, *cud* and *could* would become homophones, but *bud* would not become the same as any lexical competitor).

The carrier sentence was produced in both Sheffield and SSBE accents by the same male speaker ([əmaskɪnjəˀseɪwɜ:d---pli:s] in Sheffield and [aɪmɑ:skɪŋju:təseɪðəwɜ:d---pli:s] in SSBE). Multiple instances of the carrier sentences were recorded in each accent and one from each accent was selected for use in the experiment. The speaker had lived in Sheffield until the age of 19 and had then moved to the south of England, where he had lived for 7 years. This speaker was selected because he had an unusual ability to switch between accents at will, and was able to produce versions of both accents that sounded like those of native speakers.[2] In addition to the carrier sentences, the speaker was recorded reading a 2-min passage from a novel in both accents.

CVCs were embedded in the carrier sentences. The bursts, fricatives, and aspiration were spliced from the sentence recording, and the voiced portions were synthesized

on-line using the cascade branch of a Klatt synthesizer (Huckvale, 2003; Klatt and Klatt, 1990), to allow for fine-grained coverage of the entire vowel space. Each stimulus had a middle portion in which the formant frequencies were static, and had formant transitions appropriate for the consonants (see below). All transitions were linear. The stimuli varied in terms of F1–F3 frequencies and duration of the middle portion. F1 frequency was restricted so that it had a lower limit of 150 Hz and an upper limit of 950 Hz. F2 frequency was restricted to have a lower limit of F1+50 Hz, and had an upper limit defined by the equation

$$F2_{\text{upper-limit}} = 3000\ \text{Hz} - 1.7*F1. \quad (1)$$

F3 frequency was restricted to have a lower limit of 2000 Hz, an upper limit of 3150 Hz, and was always at least 100 Hz greater than F2. Duration of the middle portion was restricted to be greater than 20 ms and less than 403 ms.

All other synthesis parameters were chosen to mimic the natural speech recordings. For /b/-V-/d/, F1–F3 were 200, 1500, and 2400 Hz at the start of the formant transitions for /b/. The duration of the initial transition was 20 ms. F1–F3 were 200, 2300, and 3200 Hz at the end of the formant transitions for /d/. The final transition duration was 120 ms. F4 and F5 were fixed to 3200 and 4900 Hz throughout the stimulus. The bandwidths of F1–F5 were fixed to 100, 120, 150, 100, and 175 Hz. F0 (fundamental frequency) started at 116 Hz, rose to 126 Hz, and fell to 104 Hz. AV (amplitude of voicing) started at 45 dB, rose to 51 dB, and fell to 48 dB.

For /k/-V-/d/, F1–F3 began at the target formant frequencies of the vowel (i.e., there were no formant transitions at the onset of voicing, which is typical of voiceless stops). F1–F3 were 200, 1500, and 2600 Hz at the end of the formant transitions for /d/. The final transition duration was 40 ms. F4 and F5 were fixed to 3200 and 4450 Hz throughout the stimulus. The bandwidths of F1–F5 were fixed to 100, 120, 150, 150, and 175 Hz. F0 started at 125 Hz, rose to 128 Hz, and fell to 108 Hz. AV was ramped from 0 to 40 dB over the first 10 ms, rose to 50 dB, and fell to 45 dB.

For /b/-V-/θ/, F1–F3 were 200, 1300, and 2335 Hz at the start of the formant transitions for /b/. The duration of the initial transition was 20 ms. F2 and F3 were 1290 and 2400 Hz at the end of the formant transitions for [θ], and F1 ended on the target vowel frequency. The final transition duration was 20 ms. F4 and F5 were fixed to 3200 and 4900 Hz throughout the stimulus. The bandwidths of F1–F5 were fixed to 100, 160, 250, 150, and 175 Hz. F0 started at 115 Hz, rose to 125 Hz, and fell to 106 Hz. AV started at 40 dB, rose to 45 dB, and fell to 10 dB over the last 30 ms of the vowel.

After synthesis, the stimuli were processed using a multi-band filter to fine-tune the match between the long term average spectra of the synthetic and natural speech; frequencies between 0 and 1500 Hz were attenuated by 1.5 dB, frequencies between 1500 and 3500 Hz were amplified by 6 dB, and frequencies between 3500 and 5500 Hz were attenuated by 2 dB. This filter was necessary because adjustments of the Klatt synthesis parameters (e.g., formant bandwidths) were not entirely sufficient by themselves to match the voice quality of the natural speech produced by this talker.

The stimuli were put together on-line during the experiment and played at a sampling rate of 11 kHz using a computer sound card and headphones (Sennheiser HD 414) in a sound attenuated booth.

3. Procedure

There were two testing sessions, one for each accent. The order of presentation was counterbalanced across subjects. Sessions were conducted on separate days to minimize the risk that subjects would be aware that the speaker was the same in both conditions (subjects were informally questioned after completing the experiment and no subject reported that the speaker was the same). Each session was self-paced and lasted approximately 1 h. At the start of each session, subjects listened to a short passage read by the speaker to familiarize them with the accent. They then found the best exemplar for one practice word (*kid*), and best exemplars for 16 experimental words: *bad, bard, bed, bird, bud, bod, bawd, bid, bead, booed, cud, could, cooed, Beth, birth*, and *bath*. To find the best exemplars, subjects heard a synthesized word embedded in a carrier sentence on each trial, and rated whether it was close to being a good exemplar of the target word that was displayed orthographically on a computer screen. They gave their response by positioning and clicking a computer mouse on a continuous scale from *close* to *far away*. The vowel parameters (F1, F2, F3, and duration) were adjusted after each trial using a customized procedure that was inspired by computer minimization algorithms (see Press *et al.*, 1992) and was designed to find the best exemplar location for that word in this four-dimensional parameter space.

The procedure had five stages, with six trials per stage, and was able to find the best exemplar locations within this large vowel space after 30 trials. In brief, the procedure adjusted F1 and F2 in Stages 1 and 2 (starting along a path in Stage 1 that would be likely to get close to best exemplars most quickly), adjusted the more secondary dimensions of F3 and duration in Stages 3 and 4, and then fine-tuned the best exemplar location in Stage 5.

The best exemplar was found in Stage 1 along a straight-line path through the F1/F2 plane that was defined by two points: the middle of the vowel space (F1=500 Hz and F2 =1500 Hz) and the average F1 and F2 frequencies that the speaker of the carrier sentence had used for that word (averaged across the two accents). The path passed through these points and was terminated at the boundaries of the vowel space. For example, the Stage 1 search path for *bead* crossed diagonally through the vowel space, from the extreme high-front boundary of the space (i.e., low F1 and high F2), through the measured values for /i/ and the middle of the space, and through to the extreme low-back boundary of the space (i.e., high F1 and low F2). All other parameters were fixed to neutral values (F3=2500 Hz and duration=116 ms) in this stage.

On the first two trials of Stage 1, subjects heard the most extreme stimuli that it was possible to synthesize along the search path (e.g., in the case of *bead*, they heard extreme

high-front and low-back vowels, with the order of these two trials randomized). The selection of stimuli on the remaining trials was based on the subjects' judgments, using formulas that were designed to find stimuli along the path that would be perceived as better exemplars. On the third trial, subjects heard a stimulus that was selected by a weighted average of the first two stimuli, according to the equation

$$c = a^* \frac{f(b)}{f(a)+f(b)} + b^* \frac{f(a)}{f(a)+f(b)}, \tag{2}$$

where a and b are the positions on the search path for the first two trials, $f(a)$ and $f(b)$ are the goodness ratings for the stimuli on those trials (the goodness responses of *close* to *far away* were scaled from 0 to 1), and c is the new path position selected for the 3rd trial. On the 4th-6th trials, the stimuli were selected by finding the minimum of a parabola that was defined by the equation

$$\min = \frac{b - 0.5^*\{[b-a]^2*[f(b)-f(c)] - [b-c]^2*[f(b)-f(a)]\}}{[b-a]*[f(b)-f(c)] - [b-c]*[f(b)-f(a)]}, \tag{3}$$

where b was the path position of the best stimulus found thus far; a and c were the most recently tested positions on either side of b; and $f(a)$, $f(b)$, and $f(c)$ were the goodness ratings for those stimuli. In cases where Eq. (3) could not be calculated (i.e., if a, b, and c were co-linear, or b was at an extreme position on the path), a weighted average [Eq. (2)] was calculated instead, based on the best exemplar found thus far and the last stimulus that had been played. At the completion of this stage, the parameters of the best stimulus found thus far were passed onto the next stage of the search algorithm.

The same six-trial search algorithm was used for the other stages, along different paths. Stage 2 found the best exemplar along a straight-line path that was orthogonal in the F1/F2 plane to the Stage 1 path, and included the best exemplar found in Stage 1. Stage 3 searched along the F3 dimension, keeping all other parameters fixed to the best exemplar values that had been found in Stage 2. Stage 4 searched along the duration dimension (log-scaled values), keeping all other parameters fixed to the best exemplar values found in Stage 3. Stage 5 searched along a straight-line path through a three-dimensional F1, F2, and duration space (F3 did not vary), that began in the middle of the vowel space (F1=500 Hz, F2=1500 Hz, and duration=116 ms) and passed through the parameters of the best exemplar chosen thus far.

Subjects were allowed to repeat stages if they responded that the search algorithm had gone wrong (e.g., when the search was thrown off by an erroneous goodness rating). The best exemplar was defined as the stimulus given the highest goodness rating in Stage 5.

B. Results

1. Bud and cud

As displayed in Fig. 1, listeners chose different formant frequencies for *bud* and *cud* in SSBE and Sheffield carrier sentences, indicating that they normalized these vowels for accent. The shift appeared to occur predominantly along the F1 dimension; both groups of listeners chose a higher F1 for *bud* and *cud* in SSBE (Standard Southern British English) sentences than in Sheffield sentences, although the size of the shift appeared to be larger for northerners. The differences in F1 and F2 were tested in separate repeated measures ANOVA analyses, with word (*bud* or *cud*) and sentence context (SSBE or Sheffield) coded as within-subject variables, and subject background (northern or southern) coded as a between-subject variable. For F2, there were no significant main effects or interactions, $p > 0.05$, suggesting that listeners were not normalizing for accent on this dimension. For F1, however, there was a main effect of sentence context, $F(1,18) = 11.94$, $p < 0.01$, confirming that listeners overall chose higher F1 frequencies for *bud* and *cud* in the SSBE sentences.[3] There was also a main effect of subject background, $F(1,18) = 12.08$, $p < 0.01$, demonstrating that northern listeners consistently chose higher F1 values for *bud* and *cud* than did southern listeners. There was no significant main effect of word and no significant interactions for F1, $p > 0.05$.

The effects of sentence and subject background on F1 can be seen clearly in Fig. 2. In the Sheffield sentences,

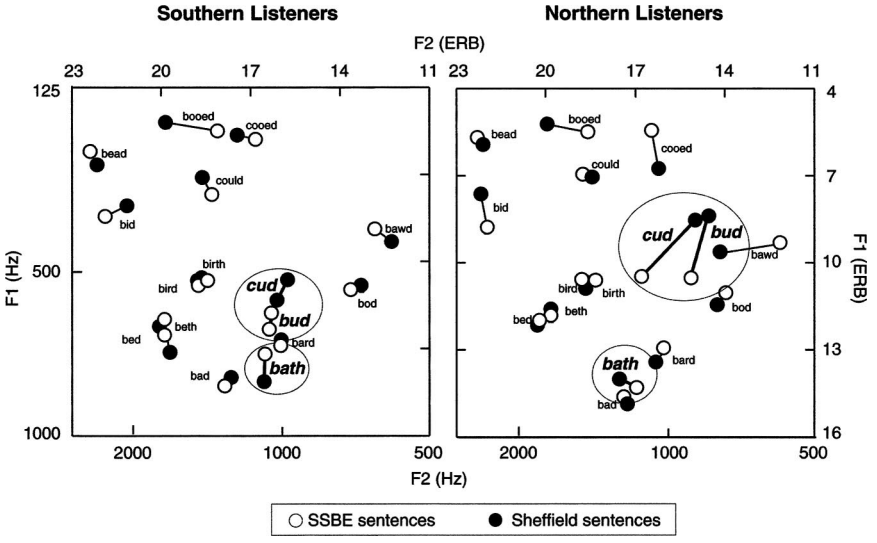


FIG. 1. Average F1 and F2 formant frequencies of best exemplars for northern and southern listeners in SSBE and Sheffield carrier sentences. The F1 frequencies of *bud* and *cud* were significantly different in the two carrier sentences for both groups of listeners, but no other words were reliably normalized for accent.

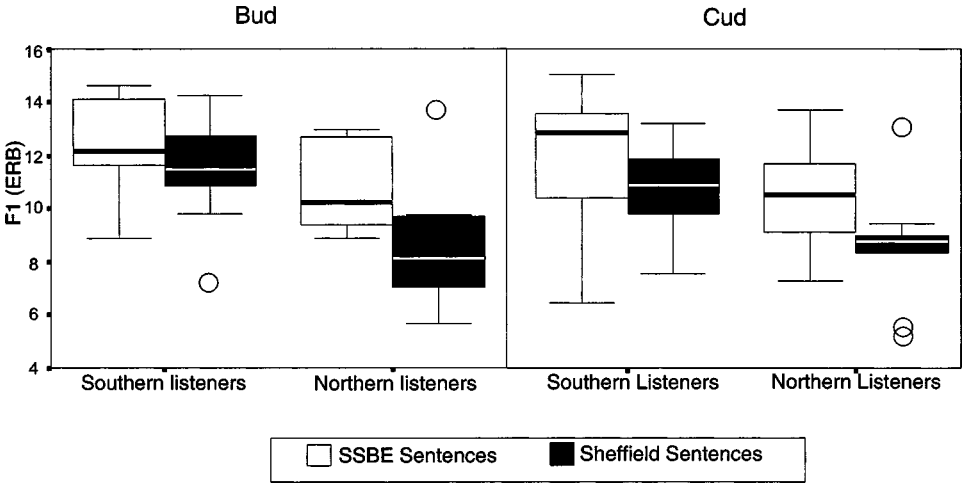


FIG. 2. Boxplots of F1 formant frequency values for *bud* and *cud* in SSBE and Sheffield carrier sentences for northern and southern listeners. Boxplots display the interquartile range of scores. The box shows the 25th to 75th percentiles, with a line at the median value. The lower and upper "whiskers" respectively show the first and last quartiles, with outliers represented by the unshaded circles. The best exemplar locations for northerners had higher F1 frequencies than those chosen by southerners, and both groups of listeners chose higher F1 frequencies in Sheffield than in SSBE carrier sentences.

northerners chose a high vowel (i.e., low F1 frequency) that was appropriate for that accent, but southerners chose a low-central vowel that was lower than Sheffield speakers actually produce. In the SSBE context, southerners chose a low vowel (i.e., high F1 frequency) that was appropriate for that accent, but northerners chose a central vowel that was higher than SSBE speakers produce. Although the size of the shift in *bud* was relatively small for southerners, the direction of this shift was consistent; nine of ten southerners chose higher F1 frequencies for *bud* in the Sheffield context.

As displayed in Tables I and II, there were few differences between *bud* and *cud* in terms of F3 or duration. Separate repeated measures ANOVA analyses for F3 and duration revealed that there were no significant main effects or interactions of sentence context, subject background, or word, $p > 0.05$, further suggesting that vowel normalization for accent only took place in the F1 dimension for *bud* and *cud*.

2. Bath

As displayed in Fig. 1, listeners chose relatively similar formant frequencies for *bath* in SSBE and Sheffield carrier sentences, with perhaps a small shift in the F1 dimension for southern listeners. Separate repeated measures ANOVA analyses for F1, F2, and F3 revealed that there were no significant main effects or interactions of sentence or subject background, $p > 0.05$, suggesting that the formant frequencies of *bath* were not consistently normalized.

As displayed in Table II, there were no strong normalization effects for duration; listeners chose similar vowel durations in both sentence contexts, although there was a trend for southerners to choose shorter vowels in the Sheffield sentences. However, there was a consistent effect of subject background; southern listeners chose a longer vowel for *bath* in both sentence contexts than did northerners. A repeated measures ANOVA analysis verified that there was a main effect of subject background, $F(1,18) = 8.09$, $p < 0.01$, but no significant main effect of sentence context or significant interactions, $p > 0.05$. The effect of subject background on duration can be seen clearly in Fig. 3. Northerners preferred shorter vowels that corresponded to their production of [a] in *bath*, and southerners preferred longer vowels that corre-

TABLE I. Average F3 frequencies (ERB) of best exemplars for northern and southern listeners in SSBE and Sheffield sentence contexts.

	Northern		Southern		
Word	SSBE	Sheffield	SSBE	Sheffield	Average
bud	23.4	23.0	22.8	23.5	23.2
cud	23.7	23.2	22.9	23.4	23.3
bath	22.7	22.1	23.1	22.1	22.5
bead	24.0	24.5	24.0	23.7	24.1
bid	24.0	24.2	23.8	23.2	23.8
bed	22.9	22.9	23.2	23.2	23.1
beth	23.4	22.4	23.2	23.4	23.1
bird	22.7	22.4	22.9	23.0	22.8
birth	22.3	22.4	23.5	23.1	22.8
bad	22.1	22.1	22.6	22.9	22.4
bard	22.8	22.9	23.6	23.2	23.1
bod	23.6	23.3	23.1	22.7	23.2
bawd	23.6	23.2	23.5	23.3	23.4
booed	22.2	22.4	22.4	22.4	22.4
cooed	22.5	22.8	22.2	23.4	22.7
could	23.3	23.0	23.4	22.8	23.1
Average	23.1	22.9	23.1	23.1	

TABLE II. Average durations (ms) of best exemplars for northern and southern listeners in SSBE and Sheffield sentence contexts.

	Northern		Southern		
Word	SSBE	Sheffield	SSBE	Sheffield	Average
bud	62.5	72.8	74.5	62.2	68.0
cud	63.7	62.0	81.8	69.7	69.3
bath	81.8	71.5	146.0	108.6	102.0
bead	138.9	129.4	120.9	130.7	130.0
bid	72.5	60.0	63.0	62.5	64.5
bed	78.1	66.8	78.1	66.8	72.5
beth	76.9	68.1	71.3	71.6	72.0
bird	182.6	144.2	119.5	130.6	144.2
birth	135.6	137.5	145.1	158.1	144.1
bad	101.0	70.6	104.6	84.7	90.2
bard	208.2	181.3	176.6	174.0	185.0
bod	68.5	79.5	77.5	74.4	75.0
bawd	187.8	170.1	176.3	151.7	171.5
booed	176.7	161.7	162.1	154.5	163.8
cooed	173.3	169.7	147.0	146.0	159.0
could	65.3	64.8	70.0	77.2	69.3
Average	117.1	106.9	113.4	107.7	

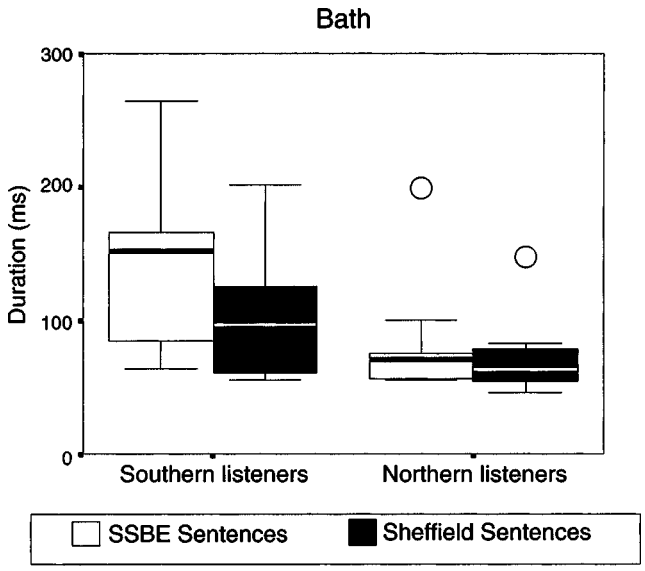


FIG. 3. Boxplots of duration values for *bath* in SSBE and Sheffield carrier sentences for northern and southern listeners. Northerners chose shorter vowels than southerners overall, but there was no normalization for the accent of the carrier sentences.

sponded to their production of [ɑ:] in *bath*. Although the formant frequencies were not significantly different, the results trended in the same direction (see Fig. 1); the median values of F1 and F2 for *bath* were more similar to *bad* ([bad]) than *bard* ([bɑ:d]) for northerners, and were more similar to *bard* than *bad* for southerners. This difference may have failed to reach significance because [a] and [ɑ:] have very similar formant frequencies overall; the vowels differ more markedly in duration.

3. Other words

As displayed in Fig. 1, listeners chose similar F1 and F2 frequencies in SSBE and Sheffield carrier sentences for most other words, with shifts for a few words such as *booed* and *bawd*. The potential differences in F1 and F2 were tested in separate repeated measures ANOVA analyses, with word (i.e., all words other than *bud*, *cud*, and *bath*) and sentence context coded as within-subject variables, and subject background coded as a between-subject variable. There was a main effect of word for F1, $F(12,216)=203.98$, $p<0.01$, and F2, $F(12,216)=113.76$, $p<0.01$, demonstrating that different words had different formant frequency values, but there were no main effects of sentence context or subject background, and no significant interactions, $p>0.05$. The differences in *bawd* and *booed* displayed in Fig. 1 were thus not reliable. As displayed in Tables I and II, listeners generally chose similar values for F3 and duration in SSBE and Sheffield sentence contexts. Separate repeated measures ANOVA analyses revealed that there was a main effect of word for F3, $F(12,216)=7.78$, $p<0.01$, and for duration, $F(12,216)=40.58$, $p<0.01$, demonstrating that different words had different F3 and duration values, but there were no main effects of sentence context or subject background, and no significant interactions, $p>0.05$.

It is notable that northern and southern listeners both chose a high-front vowel for *booed* and a high-central vowel for *cooed*, rather than high-back vowels with lower F2 frequencies (see Fig. 1). Although these preferences may seem unusual, they correspond to recent changes in the way that British English speakers produce these vowels; younger speakers in particular have begun to produce these traditionally high-back vowels with less lip rounding and with a more forward tongue position (Docherty and Foulkes, 1999; Torgersen, 1997; Williams and Kerswill, 1999).

III. EXPERIMENT 2

Experiment 1 demonstrated that individuals living in London normalized *bud* and *cud* for accent, and that the patterns of normalization depended on whether the listeners were northern or southern. Experiment 2 further examined the role of language experience on vowel normalization by testing northerners who still live in the north of England. The subjects were born and raised in Ashby de la Zouch, a market town where the dominant accent is similar to that spoken in Sheffield. The subjects were 16–17 years old, and had not yet moved for employment or university education. The aim was to determine whether the patterns of normalization found for northerners in Experiment 1 were affected by the subjects' time living in London, or whether all northerners (i.e., even those who have not lived in the south) have the same patterns of normalization.

A. Method

1. Subject selection

Twelve subjects were tested. All were native English speakers, aged 16–17 years, born and raised in Ashby de la Zouch, and reported no hearing or language problems. One subject was dropped from the experiment because her best exemplar locations were not reliable (i.e., as in Experiment 1, subjects were dropped when the best exemplar locations for the vowels that were stable between accents differed by more than 2 ERB).

2. Stimuli and apparatus

The stimuli were synthesized in advance so that the experiment could be run using a portable computer. The entire range of possible vowels was synthesized with a resolution of 0.5 ERB[4] (Moore *et al.*, 1997) in F1 and F2. Duration was quantized in 16 steps on a log scale, from 20 to 403 ms. F3 was fixed to 2500 Hz for all stimuli; although Experiment 1 had shown that F3 varied for different words, the results suggested that this parameter made only a modest contribution to perceived goodness. There were a total of 7616 stimuli synthesized for each of the CVC contexts. The stimuli and apparatus were the same as in Experiment 1 in all other respects.

3. Procedure

There was a four-stage search for best exemplars along the F1, F2, and duration dimensions, with six trials for each stage; the F3 adjustment stage was omitted. The procedure was the same as in Experiment 1 in all other respects.

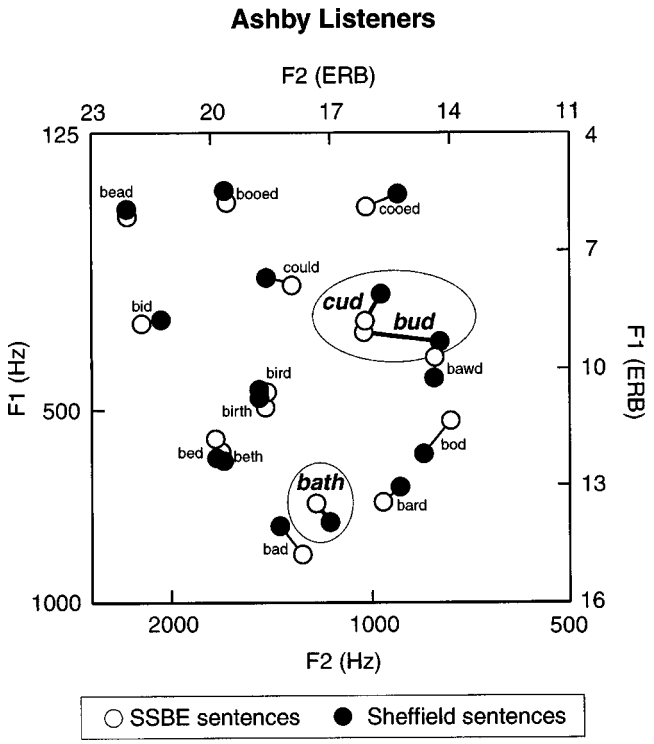


FIG. 4. Average F1 and F2 formant frequencies of best exemplar locations for Ashby listeners in SSBE and Sheffield carrier sentences. The F1 and F2 frequencies did not vary significantly between the two carrier sentences, suggesting that no vowels were normalized for accent.

B. Results

1. Bud and cud

As displayed in Fig. 4, Ashby listeners chose similar formant frequencies for *cud* in SSBE and Sheffield carrier sentences, but for *bud* there was a possible difference in the F2 dimension; listeners tended to choose a higher F2 for *bud* in SSBE than in Sheffield carrier sentences. Separate repeated measures ANOVA analyses for F1 and F2 revealed that there were no main effects of word, sentence context, or their interactions, $p>0.05$, demonstrating that the shift in the F2 dimension for *bud* was not reliable. As displayed in Table III, there was also little difference between *bud* and *cud* in terms of duration; a repeated measures ANOVA analysis revealed that there were no main effects of word, sentence context, or their interactions, $p<0.05$. There was thus no consistent evidence that Ashby listeners normalized *bud* and *cud* for accent; they chose traditionally northern vowels in both carrier sentences.

2. Bath

Ashby listeners also chose similar formant frequencies for *bath* in SSBE and Sheffield carrier sentences (see Fig. 4), indicating that there was no normalization for accent. There was a trend for listeners to choose a longer vowel in the SSBE than in the Sheffield carrier sentences (Table III). However, separate repeated measures ANOVA analyses for F1, F2, and duration revealed that there was no main effect of sentence context, $p<0.05$. There was thus no clear evidence that Ashby listeners normalized *bath* for accent.

TABLE III. Average durations (ms) of best exemplars for Ashby listeners in SSBE and Sheffield sentence contexts.

Word	SSBE	Sheffield	Average
bud	63.8	65.4	64.6
cud	76.9	55.8	66.4
bath	108.4	79.3	93.8
bead	99.0	115.7	107.4
bed	84.5	56.5	70.5
bid	71.9	68.5	70.2
beth	73.9	58.2	66.0
bird	139.9	117.3	128.6
birth	124.5	113.3	118.9
bad	85.4	77.8	81.6
bard	156.4	162.7	159.5
bod	72.4	72.0	72.2
bawd	140.1	142.5	132.3
booed	170.4	164.4	167.4
cooed	146.7	140.8	143.8
could	76.1	69.6	72.9
Average	105.6	96.4	

3. Other words

For all other words, listeners chose similar formant frequencies and durations for each target word in SSBE and Sheffield carrier sentences. Separate repeated measures ANOVA analyses for F1, F2, and duration revealed that there was a main effect of word for F1, $F(12,120)=76.08$, $p<0.01$, F2, $F(12,120)=47.24$, $p<0.01$, and duration, $F(12,120)=15.61$, $p<0.01$. However, there was no main effect of sentence context and no significant interaction with word, $p>0.05$, suggesting that none of the other words varied depending on accent.

IV. DISCUSSION

The results demonstrated that individuals living in London normalized the vowels in *bud* and *cud*—but not *bath*—for southern and northern English accents, with the patterns of normalization reflecting each listener's linguistic experience. When individuals living in London heard sentences that were similar to their native accent, they chose formant frequencies for *bud* and *cud* that matched what speakers of that accent would produce; southerners living in London selected an [ʌ] vowel (i.e., high F1) when listening to SSBE sentences and northerners living in London selected an [ʊ] vowel (i.e., low F1) when listening to Sheffield sentences. When individuals living in London heard sentences that did not match their native accent (e.g., northerners listening to SSBE speech), they chose centralized vowels for *bud* and *cud* rather than the [ʌ] and [ʊ] vowels that would normally be produced in SSBE and Sheffield accents, respectively. Northerners who were less experienced with southern accents (i.e., Ashby listeners) did not normalize for accent at all, choosing vowels in Sheffield and SSBE sentences that would be appropriate for northern speakers.

Episodic memory research has shown that individuals store phonetically detailed representations of spoken words in long-term memory (e.g., Goldinger, 1998; Nygaard and Pisoni, 1998; Palmeri *et al.*, 1993), and our working hypoth-

esis based on this research was that listeners would choose best exemplars that matched their long-term memory representations for words spoken by speakers with similar accents. It was surprising then, that northerners living in London, for example, chose best exemplars for *bud* and *cud* in the SSBE sentences that do not match how southerners actually produce these vowels [see Wells (1982) for a phonetic description of SSBE vowel production]. Although this may suggest that listeners were not performing the task using stored exemplars, it is possible that listeners were using inaccurate exemplars that had been affected by perceptual magnet effects (Iverson and Kuhl, 1995, 1996, 2000; Iverson *et al.*, 2003) or category assimilation processes (Best *et al.*, 1988, 2001; Flege, 1992, 1995). That is, the northerners' perception of SSBE [ʌ] may have been distorted because they do not have a native /ʌ/ category; northerners may perceive the SSBE [ʌ] to be a member of their native /ə/ or /ɜ:/ categories [or a new category caused by merging /ʌ/-/ʊ/; see MacKay *et al.*, (2001)], causing the SSBE [ʌ] to sound more centralized. It is plausible that such perceptual distortion caused northern listeners to remember mistakenly that southerners produce centralized vowels for *bud* and *cud*. In other words, the basic hypothesis that listeners choose best exemplars that match long-term memory representations may be correct, but the memories of listeners may be inaccurate.

There are two aspects of the present results that are inconsistent with this perceptually distorted exemplar account. First, the *bud* and *cud* vowels that southerners chose in the Sheffield sentences cannot be easily explained by perceptual magnet effects or category assimilation. Northerners produce *bud* and *cud* using [ʊ], and one would normally expect that southerners would assimilate this northern vowel into whatever native category is most similar perceptually (i.e., the southerners' own /ʊ/ category, that they use in words like *book* or *could*). Instead, southerners chose a low-central best exemplar for *bud* and *cud* that is on the other side of the vowel space from [ʊ]. It seems unlikely that southerners erroneously perceive the northern [ʊ] as a low-central vowel. Second, there was no normalization for *bath*. Southerners and northerners both use the vowels [a] and [ɑ:], and speakers of British English are very aware that the lexical distribution of these vowels is a clear marker of accent (Trudgill, 1986). Northerners in London thus know that southerners produce *bath* with [ɑ:], as do southerners in London know that northerners produce *bath* with [a]. Yet subjects in this experiment chose vowels for *bath* based on their own accent, rather than on their knowledge of what vowel would be expected based on the accent of the carrier sentence.

Although these patterns of normalization may seem idiosyncratic, they correspond closely with the changes in production that speakers tend to make when they live in multidialectal environments (Trudgill, 1986). Northerners who live in the south of England typically modify some aspects of their accent in order to fit in with southerners; they change their production of the vowel in *bud* and *cud* so that it becomes centralized (Trudgill, 1986), much like the centralized vowel that they chose as best exemplars for these words in the SSBE-accented carrier sentence. Northerners also maintain some aspects of their regional identity; they retain their [a] when producing words like *bath* [and would rather ''drop dead'' than produce these words like a southerner (Trudgill, 1986)], much like they chose [a] for *bath* in both carrier sentences. Southerners living in London, however, speak the locally dominant dialect and are less apt to modify their productions when speaking to others; they likewise made relatively small adjustments to *bud* and *cud*, and preferred southern pronunciations of *bath* in both carrier sentences.

Production may also help explain why Ashby listeners did not perceptually normalize for accent. One could imagine that Ashby listeners did not normalize because they had not had enough perceptual experience with southern accents (e.g., see Labov and Ash, 1997). However, Ashby listeners are regularly exposed to southern English accents through the media (Foulkes and Docherty, 1999) and they are able to correctly identify the accents of southern speakers in perceptual tests (Evans, 2001). Moreover, all listeners heard a short passage read by the speaker in the relevant accent before starting the experiment, and such short-term familiarization can be enough to tune speech recognition processes to the characteristics of individual talkers (Nygaard and Pisoni, 1998). It may have been more important that these listeners (aged 16–17 years, born and raised in Ashby) had not had the experience of modifying their own speech in order to fit into a new environment (e.g., when attending a university). It is thus plausible that these Ashby listeners chose northern vowels in the SSBE carrier sentence because they had not yet learned to change their speech when talking to southerners, even though they know how southerners talk (see also Flege, 2003).

The mechanism responsible for this perception-production link is unclear. The results are consistent with motor theory's claim that listeners perceive speech in terms of their own articulatory gestures (Liberman *et al.*, 1967). That is, the acquisition of new articulatory targets to modify one's own accent may have directly changed how vowels in the southern and northern accents were perceived. However, it is possible too that the best exemplars found in perceptual experiments reflect auditory targets that a listener tries to achieve when speaking (Allen and Miller, 2001). That is, listeners may need to first change their notions of which phonemes sound good, in order to learn to modify the accent of their own speech.

The current study is only a first attempt to investigate how vowel perception is modified to accommodate accent differences in the same language, but the results thus far differ from descriptions of how listeners perceive foreign or foreign-accented speech. Cross-language research has suggested that foreign-accented phonemes are assimilated into the same categories that listeners use for native speech (e.g., Best, 1994; Flege, 1992), but the present results suggest that listeners can adjust their categorizations to accommodate different accents within the same language. Cross-language research has also emphasized the age and amount of exposure as determining factors in the ability to perceive and produce a foreign language (Flege *et al.*, 1999). However, the current results suggest that perceptual adjustments for accents within the same language are not simply determined by exposure; changes in best exemplar locations appear to follow sociol-

inguistic principles that help explain what happens to speech when an individual chooses to fit in with a particular community or subculture. It is likely that these sociolinguistic principles [e.g., a listener's motivation to learn; see Piske *et al.* (2001)] also have a role in cross-language cases. That is, although learning to perceive and produce the phonemes of a new language must depend on exposure to a great extent, losing one's accent may also be affected by one's willingness to be identified as a member of the same culture as a native speaker of that language.

ACKNOWLEDGMENTS

This research was supported by an EPSRC Doctoral Training Award. We are grateful to Richard Dewire and Alexandra Evans for their assistance with the experiments.

[1]A broad definition of *normalization* is used here (i.e., the perceptual and cognitive adjustments that allow listeners to accommodate differences between speakers), rather than the narrow definition that is sometimes used [i.e., a hypothetical perceptual process in which speaker-specific information is discarded (Pisoni, 1997)].

[2]The speaker was coached to produce words like *bath* and *ask* with an [ɑː] in SSBE sentence context, because he normally produced these words with [a].

[3]Note that there were no main effects of sentence context for the other vowels. Thus, the differences in the sentence contexts did not cause listeners to shift the overall vowel space due to some experimental artifact; the listeners specifically normalized *bud* and *cud*.

[4]Even though the quantization of the stimulus space used ERB values, the algorithm searched through the vowel space along linear Hz paths. The linear Hz values were rounded in the search algorithm to the values of the nearest available stimulus. This allowed the search algorithm to be the same as in experiment 1, and still have a stimulus quantization that was perceptually uniform.

Allen, J. S., and Miller, J. L. (**2001**). ''Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate,'' Percept. Psychophys. **63**, 798–810.

Best, C. T. (**1994**). ''The Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimilation Model,'' in *The Development of Speech Perception*, edited by J. C. Goodman and H. C. Nusbaum (MIT, Cambridge, MA), pp. 167–244.

Best, C. T., McRoberts, G. W., and Goodell, E. (**2001**). ''Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system,'' J. Acoust. Soc. Am. **109**, 775–794.

Best, C. T., McRoberts, G. W., and Sithole, N. N. (**1988**). ''The phonological basis of perceptual loss for nonnative contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants,'' Human Percept. Perform. **14**, 345–360.

Bohn, O.-S., and Flege, J. E. (**1993**). ''Perceptual switching in Spanish/English bilinguals,'' J. Phonetics **21**, 267–290.

Clarke, C. M. (**2002**). ''Perceptual Adjustment to Foreign-Accented English With Short Term Exposure,'' paper presented at the 7th International Conference on Spoken Language Processing (ICSLP), Denver, CO.

Docherty, G. J., and Foulkes, P. (**1999**). ''Derby and Newcastle: Instrumental phonetics and variationist studies,'' in *Urban Voices: Accent Studies in the British Isles*, edited by P. Foulkes and G. J. Docherty (Arnold, London), pp. 47–71.

Elman, J. L., Diehl, R. L., and Buchwald, S. E. (**1977**). ''Perceptual switching in bilinguals,'' J. Acoust. Soc. Am. **62**, 971–974.

Evans, B. G. (**2001**). ''An investigation into the identification of four varieties of English,'' unpublished master's thesis. University of Newcastle upon Tyne, Newcastle upon Tyne, U.K.

Flege, J. E. (**1991**). ''Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language,'' J. Acoust. Soc. Am. **89**, 395–411.

Flege, J. E. (**1992**). ''The intelligibility of English vowels spoken by British and Dutch talkers,'' in *Intelligibility in Speech Disorders: Theory, Measurement and Management*, edited by R. D. Kent (Benjamins, Amsterdam), pp. 157–232.

Flege, J. E. (**1995**). ''Second language speech learning: Theory, findings, and problems,'' in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 233–277.

Flege, J. E. (**2003**). ''Assessing constraints on second-language segmental production and perception,'' in *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, edited by A. Meyer and Niels Schiller (Mouton de Gruyter, Berlin).

Flege, J. E., and Eefting, W. (**1987**). ''Cross-language switching in stop consonant perception and production by Dutch speakers of English,'' Speech Commun. **6**, 185–202.

Flege, J. E., Mackay, I. R. A., and Meador, D. (**1999**). ''Native Italian speakers' perception and production of English vowels,'' J. Acoust. Soc. Am. **106**, 2973–2987.

Foulkes, P., and Docherty, G. J. (**1999**). ''Urban Voices—Overview,'' in *Urban Voices: Accent Studies in the British Isles*, edited by P. Foulkes and G. J. Docherty (Arnold, London), pp. 1–24.

Goldinger, S. D. (**1996**). ''Words and Voices: Episodic Traces in Spoken Word Identification and Recognition Memory,'' Exp. Physiol. **22**, 1166–1183.

Goldinger, S. D. (**1998**). ''Echoes of Echoes? An Episodic Theory of Lexical Access,'' Psychol. Rev. **105**, 251–279.

Hazan, V. L., and Boulakia, G. (**1992**). ''Perception and Production of a voicing contrast by French-English bilinguals,'' Lang Speech **36**, 17–38.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

Huckvale, M. (**2003**). Speech Filing System (computer software) (University College, London).

Iverson, P., and Kuhl, P. K. (**1995**). ''Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling,'' J. Acoust. Soc. Am. **97**, 553–562.

Iverson, P., and Kuhl, P. K. (**1996**). ''Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/,'' J. Acoust. Soc. Am. **99**, 1130–1140.

Iverson, P., and Kuhl, P. K. (**2000**). ''Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism?'' Percept. Psychophys. **62**, 874–883.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (**2003**). ''A perceptual interference account of acquisition difficulties for non-native phonemes,'' Cognition **87**, B47–B57.

Johnson, K. (**1997**). ''Speech Perception without Speaker Normalization: An Exemplar Model,'' in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullenix (Academic, San Diego), pp. 145–165.

Klatt, D. H., and Klatt, L. C. (**1990**). ''Analysis, synthesis, and perception of voice quality variations among female and male talkers,'' J. Acoust. Soc. Am. **87**, 820–857.

Labov, B., and Ash, S. (**1997**). ''Understanding Birmingham,'' in *Language Variety in the South Revisited*, edited by C. Bernstein, T. Nunally, and R. Sabino (Alabama, Tuscaloosa).

Ladefoged, P., and Broadbent, D. (**1957**). ''Information conveyed by vowels,'' J. Acoust. Soc. Am. **29**, 98–104.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). ''Perception of the speech code,'' Psychol. Rev. **74**, 431–461.

MacKay, I. R. A., Flege, J. E., Piske, T., and Schirru, C. (**2001**). ''Category restructuring during second language acquisition,'' J. Acoust. Soc. Am. **110**, 516–528.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). ''A Model for the Prediction of Thresholds, Loudness, and Partial Loudness,'' J. Audio Eng. Soc. **45**, 224–240.

Nearey, T. M. (**1989**). ''Static, dynamic, and relational properties in vowel perception,'' J. Acoust. Soc. Am. **85**, 2088–2113.

Nygaard, L. C., and Pisoni, D. B. (**1998**). ''Talker-specific Learning in Speech Perception,'' Percept. Psychophys. **60**, 355–376.

Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. (**1993**). ''Episodic encoding of voice attributes and recognition memory for spoken words,'' J. Exp. Psychol. Learn. Mem. Cogn. **19**, 309–328.

Piske, T., Mackay, I. R. A., and Flege, J. E. (**2001**). ''Factors affecting degree of foreign accent in an L2: A review,'' Phonetics **29**, 191–215.

Pisoni, D. B. (**1997**). ''Some Thoughts on 'Normalization,' in Speech Perception,'' in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullenix (Academic, San Diego), pp. 9–32.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. H. (**1992**). *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge U.P., Cambridge).

Torgersen, E. (**1997**). ''Some phonological innovations in south-eastern British English,'' unpublished master's thesis, University of Bergen, Bergen, Norway.

Trudgill, P. (**1986**). *Dialects in Contact* (Basil Blackwell, Oxford).

Upton, C., and Widdowson, J. D. A. (**1996**). *An Atlas of English Dialects* (Oxford U.P., Oxford).

Watt, D. J. L. (**1998**). ''Variation and Change in the Vowel System of Tyneside English,'' unpublished Ph.D. dissertation, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne, U.K.

Wells, J. C. (**1982**). *Accents of English* (Vol. 2). (Cambridge U.P., Cambridge).

Williams, A., and Kerswill, P. (**1999**). ''Dialect levelling: Change and continuity in Milton Keynes, Reading and Hull,'' in *Urban Voices: Accent Studies in the British Isles*, edited by P. Foulkes and G. J. Docherty (Arnold, London), pp. 141–162.

Acoustic properties of naturally produced clear speech at normal speaking rates

Jean C. Krause[a] and Louis D. Braida
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 10 December 2002; revised 31 October 2002; accepted 3 November 2003)

Sentences spoken ''clearly'' are significantly more intelligible than those spoken ''conversationally'' for hearing-impaired listeners in a variety of backgrounds [Picheny *et al.*, J. Speech Hear. Res. **28**, 96–103 (1985); Uchanski *et al.*, *ibid.* **39**, 494–509 (1996); Payton *et al.*, J. Acoust. Soc. Am. **95**, 1581–1592 (1994)]. While producing clear speech, however, talkers often reduce their speaking rate significantly [Picheny *et al.*, J. Speech Hear. Res. **29**, 434–446 (1986); Uchanski *et al.*, *ibid.* **39**, 494–509 (1996)]. Yet speaking slowly is not solely responsible for the intelligibility benefit of clear speech (over conversational speech), since a recent study [Krause and Braida, J. Acoust. Soc. Am. **112**, 2165–2172 (2002)] showed that talkers can produce clear speech at normal rates with training. This finding suggests that clear speech has inherent acoustic properties, independent of rate, that contribute to improved intelligibility. Identifying these acoustic properties could lead to improved signal processing schemes for hearing aids. To gain insight into these acoustical properties, conversational and clear speech produced at normal speaking rates were analyzed at three levels of detail (global, phonological, and phonetic). Although results suggest that talkers may have employed different strategies to achieve clear speech at normal rates, two global-level properties were identified that appear likely to be linked to the improvements in intelligibility provided by clear/normal speech: increased energy in the 1000–3000-Hz range of long-term spectra and increased modulation depth of low frequency modulations of the intensity envelope. Other phonological and phonetic differences associated with clear/normal speech include changes in (1) frequency of stop burst releases, (2) VOT of word-initial voiceless stop consonants, and (3) short-term vowel spectra.
 [DOI: 10.1121/1.1635842]

PACS numbers: 43.71.Ky, 43.66.Ts, 43.70.Gr [KWG] Pages: 362–378

I. INTRODUCTION

When confronted with difficult communication environments, many speakers adopt a speaking style which permits them to be understood more easily. Previous studies have demonstrated that this speaking style, known as *clear speech*, is significantly more intelligible than conversational speech for hearing-impaired listeners in a quiet background as well as for both normal-hearing and hearing-impaired listeners in noise and reverberation backgrounds (Picheny *et al.*, 1985; Uchanski *et al.*, 1996; Payton *et al.*, 1994). Furthermore, the intelligibility advantage is independent of listener, presentation level, and frequency-gain characteristic (Picheny *et al.*, 1985). These results suggest that signal processing schemes that convert conversational speech to a sufficiently close approximation of clear speech could improve speech intelligibility in many situations.

In order to develop such signal processing schemes, however, it is first necessary to identify the acoustical factors responsible for the high intelligibility of clear speech. Identifying these characteristics has proven difficult because the typical speaking rate for clear speech (100 words per minute) is roughly half that of conversational speech (Picheny *et al.*, 1986). Therefore, many of the acoustical differences between clear speech and conversational speech that have been described previously (Picheny *et al.*, 1986) (e.g., more frequent and longer pauses) are likely to be related to differences in rate. A recent study (Krause and Braida, 2002), however, has demonstrated that with training, talkers can produce clear speech at normal speaking rates. In that study, clear and conversational speech were obtained from five talkers at three different relative speaking rates: slow, normal, and quick, resulting in six elicited speaking styles (referred to by mode/rate): clear/slow, conv/slow, clear/normal, conv/normal, clear/quick, conv/quick. As shown in Fig. 1, intelligibility tests of normal-hearing listeners with simulated hearing losses (achieved by additive noise) showed that all five talkers were able to produce a form of clear speech at normal rates (within 25 wpm of conv/normal speech) that provided an intelligibility advantage of at least 5 percentage points over conv/normal speech (T1 met this criterion with clear/quick speech, which was produced within 25 wpm of her conv/normal speech). Thus, speaking slowly is not solely responsible for the intelligibility benefit of clear speech. Moreover, the intelligibility advantage of clear/normal over conv/normal speech of key words in nonsense sentences (14 percentage points on average) was comparable to the that of clear/slow over conv/slow (12 points) for nearly all talkers and listeners, demonstrating that the relative intelligibility benefit of clear speech can be largely independent of rate (for a range of speaking rates, when talkers are provided with training), as well as talker and listener.

It is worth noting that this finding is not at all equivalent

[a]Present address: Department of Communication Sciences and Disorders, University of South Florida, Tampa, FL. Electronic mail: jkrause@cas.usf.edu

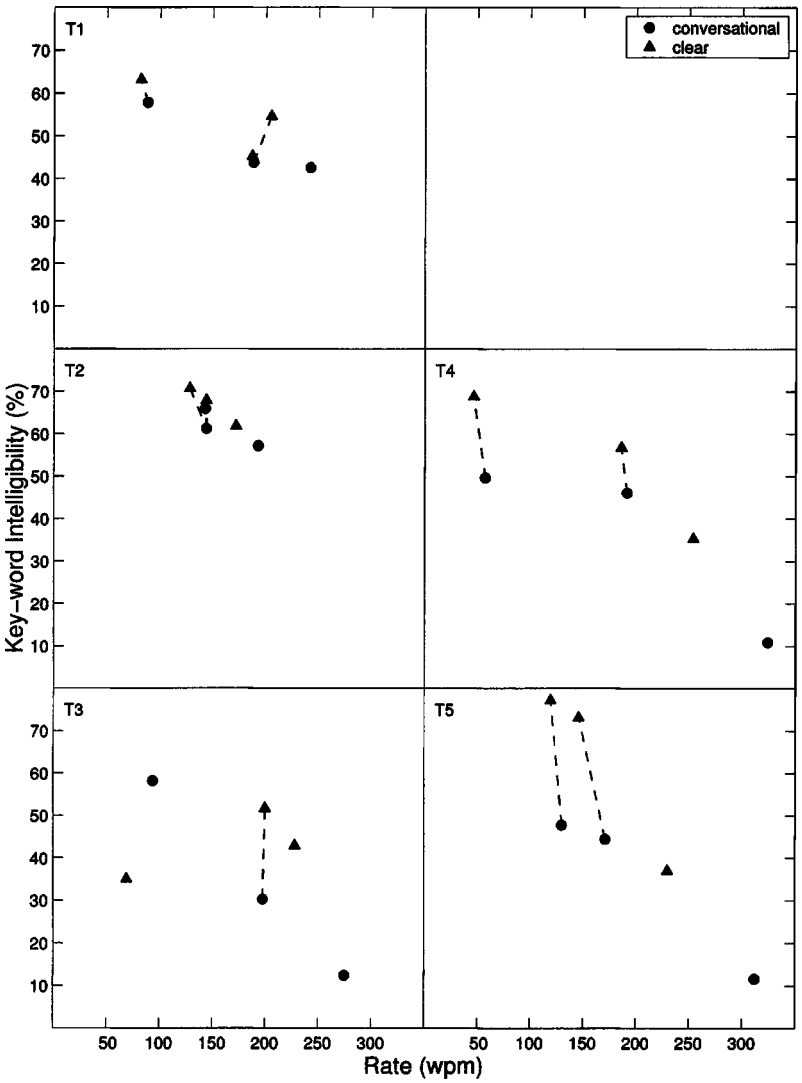


FIG. 1. Key-word scores for each talker, averaged over listener, versus speaking rate. Regardless of the speaking rate intended by the talker, dashed lines represent instances where intelligibility was improved (at least 5 percentage points) without a significant change (no more than 25 wpm) in actual speaking rate. That is, T1 achieved clear speech at normal rates when attempting to produce clear/quick speech, since her speaking rate was within 25 wpm of that for conv/normal speech. Two dashed lines emanate from T2's conv/normal data point, because both clear/slow and clear/normal speech provided an intelligibility advantage without a significant change in rate.

to demonstrating that intelligibility itself is independent of rate. Indeed, speaking slowly improved intelligibility of both conversational speech and clear speech. Yet, the *benefit* of speaking clearly remained constant (12–14 points) for speaking rates ranging from slow through normal. As such, it seems likely that clear speech has some inherent acoustic properties, independent of rate, that account for its higher intelligibility.

This paper attempts to identify one or more of these acoustic properties by directly comparing the acoustic properties of clear and conversational speech at the same rate. Although similar acoustic analyses have been made previously (Picheny *et al.*, 1986; Uchanski *et al.*, 1996), such studies were limited to clear speech that was produced at slower speaking rates than conversational speech. Differences that have been reported include the tendency of first and second formant frequencies of tense vowels to approach target values and to cluster more tightly in clear speech than in conversational speech (Chen, 1980). This result is supported by the preference of listeners for hyperarticulated vowels, or vowels with more extreme values of first and second formants, in certain speech perception tasks (Johnson, 1993). Other reported differences in clear speech relative to conversational speech include various phonological phenomena, changes in consonant power, more frequent and longer pauses, and an increase in fundamental frequency average and range (Picheny *et al.*, 1986) as well as more salient word boundary cues (Cutler and Butterfield, 1990, 1991). While some of these differences may be related to the intelligibility difference between the two speaking modes, other differences are likely to be related to differences in rate. Therefore, analyzing the acoustic differences that remain between clear and conversational speech spoken at the same rate should lend insight into the acoustic properties specifically responsible for the relative intelligibility advantage of clear speech. Toward this end, this paper presents a detailed acoustic analysis of clear/normal and conv/normal speech. The results of these analyses could ultimately lead to better signal processing approaches for digital hearing aids.

II. ACOUSTICAL ANALYSIS

The database for the acoustical analysis was the speech that was obtained in our most recent study (Krause and Braida, 2002), in which talkers were trained to produce clear speech at normal rates. All of the speech elicited from the five talkers in that study was available for analysis. The speech consisted of sentence lists (50 sentences each) from

TABLE I. Summary of acoustic measurements conducted for each talker and list. N1–N5, M1–M5, and Q1 represent unique lists of 50 sentences. With the exception of T1, the same set of 50 sentences spoken by the same talker in two different speaking styles was available for analysis. Phonological and phonetic level measurements were limited to T4 and T5. Global (sentence-level) measurements were conducted on all talkers for all speaking styles.

	Measurements—normal rate lists		Measurements—mixed rate lists	
Talker	List	Speaking style	List	Speaking style
T1		Global level only		Global level only
	N1	conv/normal	M1	conv/normal
	Q1	clear/normal	M1	clear/slow
T2		Global level only		Global level only
	N2	conv/normal	M2	conv/normal
	N2	clear/normal	M2	clear/slow
T3		Global level only		Global level only
	N3	conv/normal	M3	conv/normal
	N3	clear/normal	M3	clear/slow
T4		Global, phonological, phonetic		Global, phonological, phonetic[a]
	N4	conv/normal	M4	conv/normal
	N4	clear/normal	M4	clear/slow
T5		Global, phonological, phonetic		Global, phonological, phonetic[a]
	N5	conv/normal	M5	conv/normal
	N5	clear/normal	M5	clear/slow

[a]Automated measurements only.

the corpus of nonsense sentences described by Picheny *et al.* (1985). These sentences have five to eight words each and provide no semantic context that could aid listeners in identifying key words, which are defined as all nouns, verbs, and adjectives (e.g., ''The right cane could guard an edge.'').

The portion of the database pertaining to clear and conversational speech consisted of 300 utterances per talker, or three unique sentence lists recorded in two speaking styles. In particular, for each talker, one unique list of 50 sentences was available for analysis in both conv/normal and clear/normal modes, another list was available in conv/normal and clear/slow modes, and a third list was available in conv/slow and clear/slow modes. In the case of T1, the 50 sentences originally elicited as clear/quick speech were substituted for clear/normal speech, because these sentences met the criteria for clear speech at normal rates (produced within 25 wpm of the normal rate and providing an intelligibility benefit over conv/normal speech of at least five percentage points) and her clear/normal speech did not. Thus, with the exception of T1, the exact same sentences spoken in clear and conversational modes could be directly compared for each talker, without speaking rate as a factor.

In order to facilitate comparisons with previous studies on the acoustic properties of clear/slow speech, the acoustical analysis comprised a series of measurements that paralleled those made previously by Picheny *et al.* (1986) and Uchanski *et al.* (1996) for clear/slow speech. As in those studies, acoustical measurements were conducted at three levels of detail: global (sentence-level), phonological, and phonetic. While these categories are certainly not independent, they are convenient for structuring an initial analysis and discussion of the acoustic data as well as for making comparisons with data obtained previously for clear/slow speech.

A summary of the measurements that were conducted for each talker and list is shown in Table I. Due to the laborious and time-intensive nature of some of the measurements and of the phonetic labelling procedure, not all measurements were made on all of the available stimuli. In general, measurements for each talker were confined to 100 utterances, i.e., the 50 sentences recorded by each talker in both conv/normal and clear/normal modes. In addition, any measurements that could be automated (e.g., pause distribution) were also conducted on the 50 sentences recorded by each talker in conv/normal and clear/slow styles, as an attempt to replicate some of the previous data on the acoustic properties of clear/slow speech. Lastly, analysis at the phonological and phonetic level was limited to two talkers, T4 and T5. While analysis of all five talkers would be ideal, identifying acoustic properties of even one talker's clear/normal speech should increase our understanding of highly intelligible speech. These talkers were selected for this initial analysis based on their ability to manipulate both rate and clarity with more accuracy than the other talkers (see Fig. 1—these talkers achieved intelligibility improvements at both slow and normal rates by speaking clearly, and the improvements were larger than those achieved by T1 and T2).

A. Global measurements

To characterize properties at the sentence level, differences in pause structure, fundamental frequency distribution, and long-term rms spectra were measured, as in Picheny *et al.* (1986). In addition, temporal envelope modulations were examined, since recent studies (Shannon *et al.*, 1995) suggest that these modulations play an important role in speech intelligibility.

TABLE II. Mean values of $F0\_AVG$ and $F0\_RANGE$ for conv/normal, clear/normal, and clear/slow speech. Diff column shows only those differences that were significant in paired *t*-tests at the $p=0.05$ level.

		Normal rate lists			Mixed rate lists		
Talker		Conv/norm (Hz)	Clr/norm (Hz)	Diff (semitones)	Conv/norm (Hz)	Clr/norm (Hz)	Diff (semitones)
T1	Avg	198	202	⋯	197	211	+1.24
	Range	197	207	⋯	194	220	+2.16
T2	Avg	202	241	+3.00	203	218	+1.19
	Range	246	241	⋯	239	238	⋯
T3	Avg	197	166	−3.03	196	189	−0.64
	Range	152	128	−3.02	160	218	+5.41
T4	Avg	216	210	−0.50	212	196	−1.30
	Range	201	164	−3.54	207	214	⋯
T5	Avg	105	138	+4.63	102	153	+6.97
	Range	77	107	+5.61	69	160	+14.66

1. Pause length distribution

Pause frequency and duration were measured for talkers in conv/normal, clear/normal, and clear/slow speaking modes. Similar to Picheny *et al.* (1986), pauses were defined to be any period of silence of at least 10 ms in duration.

As in previous work (Picheny *et al.*, 1986), a dramatic increase in pause frequency and duration for clear/slow speech was measured relative to conv/normal speech: averaged across talkers, a set of 50 conv/normal sentences had 318 pauses averaging 42 ms in length while a set of 50 clear/slow sentences had 632 pauses averaging 130 ms in length. However, the pause distribution for clear/normal speech (mean pause length: 49 ms, mean number of pauses per 50 sentences: 339) was nearly the same as that of conv/normal speech. Thus, when constraints are placed on the talker's rate, the extraneous pauses found in clear/slow speech are largely eliminated. Consequently, it appears that increases in pause duration or frequency are not necessary components of highly intelligible speech.

2. Fundamental frequency distribution

Using the pitch estimation program provided in the *ESPS/waves+* software package, fundamental frequency (F0) values were extracted at 10-ms intervals from voiced portions of the speech signal. For each sentence, the average F0 value and range of F0 values attained were both then calculated. Statistical comparisons of the F0 average and range for conv/normal and clear/normal sentences as well as for conv/normal and clear/slow sentences were performed via paired *t*-tests. The difference variables were $D=F0\_AVG_{clear}-F0\_AVG_{conv}$ and $D=F0\_RANGE_{clear}-F0\_RANGE_{conv}$, where $F0\_AVG$ represents the average F0 value attained in a sentence, $F0\_RANGE$ represents the range of F0 values attained in a sentence, and subscripts indicate the speaking mode.

F0 averages and ranges as well as any corresponding difference variables that were statistically significant ($p=0.05$) are listed in Table II. Although differences in F0 average were significant for most talkers, most differences were 3 semitones or less. The notable exception, however, is T5, whose F0 average increased by 6.97 semitones for clear/slow speech relative to conv/normal speech and 4.63 semitones for clear/normal speech relative to conv/normal speech. Similarly, this talker's F0 range increased by 14.66 semitones for clear/slow speech relative to conv/normal speech and 5.61 semitones for clear/normal speech relative to conv/normal speech. This increase in F0 average and range is consistent with values previously reported for clear/slow speech (Picheny *et al.*, 1986). Yet only one other talker (T3) exhibited such a substantial increase (4 semitones or greater) in either F0 average or range in clear/slow speech, increasing range by 5.41 semitones, and no other talker exhibited such an increase for clear/normal speech. In fact, two talkers (T3 and T4) exhibited decreases in F0 average and range for clear/normal speech relative to conv/normal speech. Thus, T5 and, to a lesser extent, T3, were the only talkers to exhibit a change in F0 behavior that was consistent with previous studies of clear/slow speech (Picheny *et al.*, 1986; Bradlow *et al.*, 2003). It is possible that gender differences may explain some of the differences in F0 behavior between talkers, since T5 is the only male talker in this study, and four of the five talkers that have been examined in previous studies were also male (Picheny *et al.*, 1986; Bradlow *et al.*, 2003). However, more talkers of both genders must be examined before any conclusions can be drawn.

In addition to the statistical comparisons of F0 in each speaking style, a histogram of the F0 values and a crude estimate of sentence level intonation contours were calculated to facilitate comparisons with previous measures of conv/normal and clear/slow speech by Picheny *et al.* (1986). The F0 contours were obtained by measuring F0 at two points in each sentence, at its maximum value and 50 ms before the end of the sentence, and these values were averaged over speaking style. Figure 2 shows the results for T5, the talker with the largest amount of variation between speaking styles. The F0 distributions and corresponding F0 contours for each speaking mode reflect the increase in F0 average and range observed for this talker in both clear/slow and clear/normal speech. In particular, the F0 contours show that the maximum F0 value is larger for clear/normal speech relative to conv/normal speech, and the value for clear/slow speech is even larger. However, there is little change in the sentence-final pitch across speaking styles. Again, T5's F0 behavior for clear/slow speech (both distribution and con-

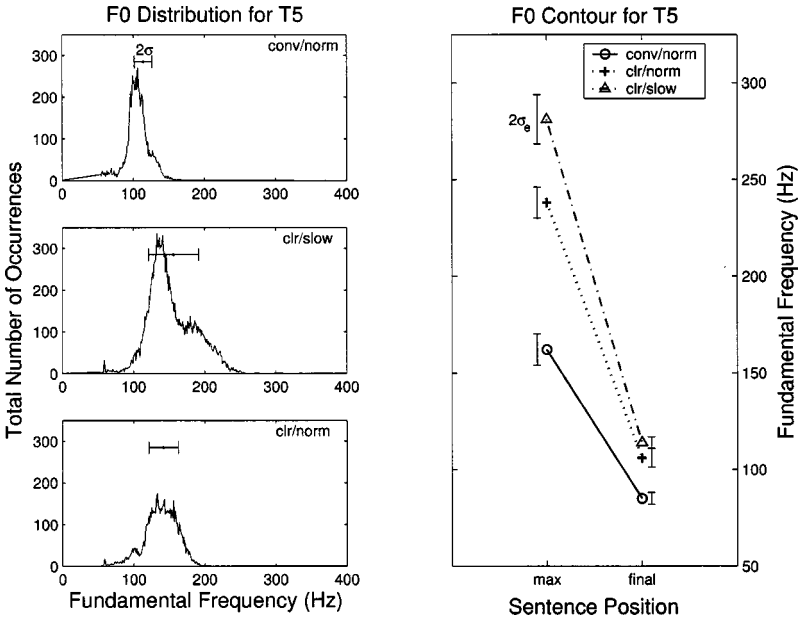


FIG. 2. Fundamental frequency (F0) data for T5, the talker who exhibited differences between speaking modes. The three graphs at left show distributions for each speaking mode; mean and standard deviation give indication of average and range. The graph at right shows the maximum F0 attained followed by the value 50 ms before the end of the sentence, on average for each speaking mode (a rough estimate of F0 contour); error bars indicate the standard error for each data point.

tour) is consistent with that previously reported by Picheny *et al.* (1986).

3. Long-term spectra

The long-term spectra of conv/normal, clear/normal, and clear/slow speech, normalized for long-term rms level, were computed using 25.6-ms nonoverlapping Hamming windows. A $\frac{1}{3}$-octave representation of the average magnitude spectra was obtained by summing components over $\frac{1}{3}$-octave intervals with center frequencies ranging from 62.5 to 8000 Hz. In addition, spectral differences were obtained by computing the ratio of the clear/normal and clear/slow spectra to the conv/normal spectrum, in order to examine the relative distribution of spectral energy for the clear modes relative to conv/normal speech. These relative spectra and the absolute spectra from which they were derived are presented in Fig. 3.

For all talkers except T1, when clear/normal and clear/slow speech provide an intelligibility advantage (all instances other than T3's clear/slow speech, see Fig. 1), the corresponding long-term spectrum has relatively more energy above 1 kHz than conversational speech. To quantify the relationship between long-term spectrum changes and intelligibility changes, the speech intelligibility index, or SII (ANSI-S3.5-1997, 1997), was computed for each condition from the third-octave band levels for that condition as well as the third-octave band levels of the additive noise used in the intelligibility experiments. The results of the SII calculations and corresponding intelligibility scores, from Krause and Braida (2002), are displayed in Fig. 4. The SII, a measure of how much speech information is audible to listeners, clearly differentiates the relative intelligibility of talkers and modes. It is highly correlated with the measured intelligibility scores, with a correlation coefficient of 0.82. Since the SII is calculated from third-octave band levels, it follows that the differences in the long-term spectra between conversational and clear speech are significant. Moreover, the SII calculation weights most heavily the audibility of speech information in the 1000–3150-Hz third-octave bands. Consequently, the relative increase of spectral energy in this frequency range, observed for all instances of clear speech that provided an intelligibility advantage for talkers T2–T5, is largely responsible for the corresponding increase in SII. To the extent that the SII is correlated with intelligibility scores, a relative increase in spectral energy in the 1000–3150-Hz range must therefore contribute to increased intelligibility.

A similar change in long-term spectrum for clear/slow speech relative to conv/normal speech was reported by Picheny *et al.* (1986) for only one of three talkers and was not considered to be ''substantial.'' However, the result is typical of a more pressed voice due to increases in vocal effort and vocal level, which are exhibited in the production of clear speech (Picheny *et al.*, 1986). While an increase in vocal level has been shown to increase middle- and high-frequency components of the spectrum relative to low-frequency components (Licklider *et al.*, 1955), it has not been shown to produce large improvements in intelligibility (Pickett, 1956). Consequently, the high-frequency emphasis alone cannot explain the large intelligibility advantage of clear/normal speech over conversational speech. It may, however, be a contributing factor, since information contained in the high-frequencies is important for cueing place of articulation (Miller and Nicely, 1955).

4. Temporal envelope modulations

Because temporal envelope modulations are thought to be important for cueing manner and voicing (Shannon *et al.*, 1995), the spectra of the octave band envelope modulations were examined for clear/slow, clear/normal, and conv/normal speech. To compute the envelope spectra, all 50 sentence files in each condition were first concatenated to mimic running speech. Because the sentence files were not only padded with 0.5 s of leading and trailing silence but also included the natural variations of silence at the beginning and ending of the sentence recordings, no attempt was made to vary intersentence duration further in concatentation. For normal rate speaking styles (conv/normal and clear/normal), the

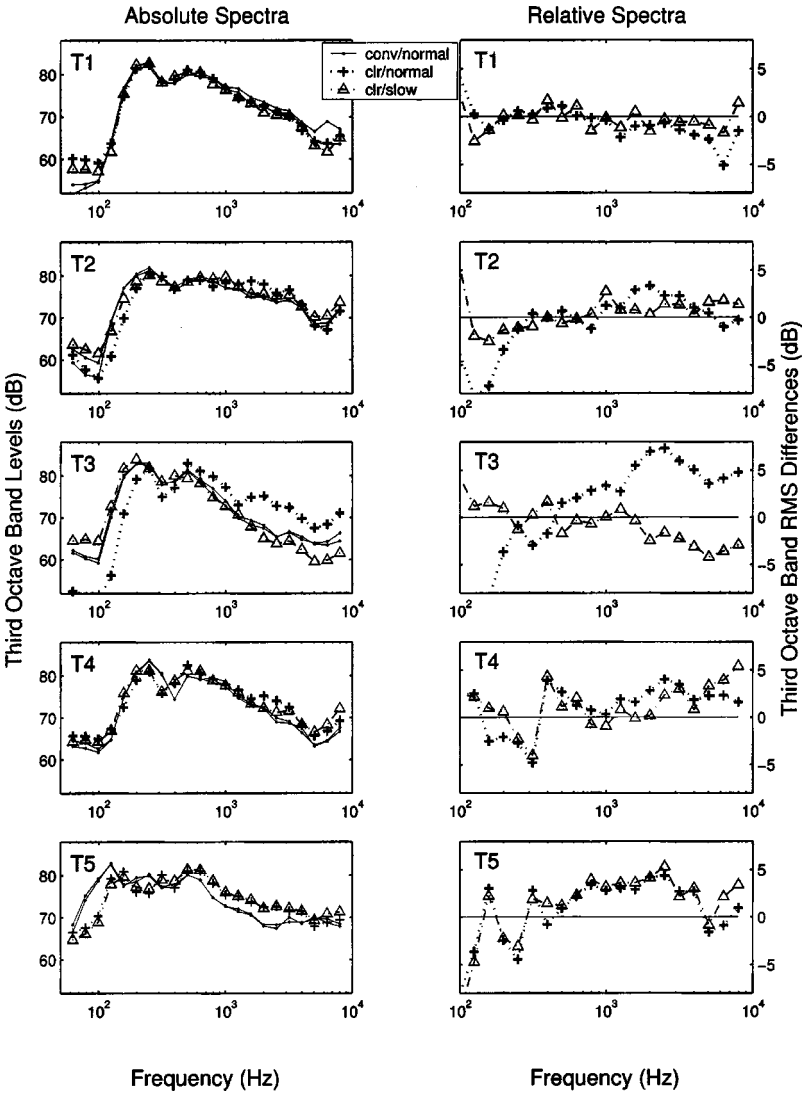


FIG. 3. Third-octave band rms data. Graphs at left show absolute spectra for each talker. Corresponding graphs at right show spectral differences, obtained by subtracting the conv/normal spectrum from the clear/normal and clear/slow spectra, which depicts the relative distribution of spectral energy between conversational and clear speech.

length of sentence files (including leading and trailing periods of silence) varied naturally between 1.7 and 3.3 s. Thus, any sentence-length artifacts in the modulation spectra of conv/normal and clear/normal speech that remained despite this natural variation in length would be confined to modulation frequencies below 0.6 Hz (1/1.7 s). Similarly, no sentence-length artifacts would be expected for clear/slow speech, since all clear/slow files were substantially longer in duration than 2.5 s, the period of the lowest modulation frequency (0.4 Hz) that was examined.

After concatenation, the speech was filtered into seven component signals, using a bank of fourth-order octave-bandwidth Butterworth filters, with center frequencies 125–8000 Hz (since the sentences were digitized at a 20-kHz sampling rate, the 8000-Hz filter was implemented as a high-pass filter). The filter bank outputs for each of the seven octave bands were then squared and low-pass filtered by an eighth order Butterworth filter with a 60-Hz cutoff frequency in order to obtain relatively smooth intensity envelopes. Finally, the intensity envelopes were downsampled by a factor of 100, and power spectra were computed. A $\frac{1}{3}$-octave representation of the spectra was obtained by summing components over $\frac{1}{3}$-octave intervals with center frequencies ranging from 0.4 to 20 Hz. As in Houtgast and Steeneken (Houtgast and Steeneken, 1985), the power spectra were normalized by the mean of the envelope function. For a single 100% modulated sine-wave, this method would result in a value of 1.0 for the $\frac{1}{3}$-octave band corresponding to the modulation frequency and zero for the other bands. In other words, to the extent that only one modulation exists per $\frac{1}{3}$-octave band, each normalized value can be considered the modulation index, *m*, for that band. The modulation index measures depth of modulation and is defined as the ratio K/A, where K denotes the maximum amplitude of the sinusoidal component of a signal, and A represents its dc component.

Figure 5 shows the spectra of the intensity envelopes and corresponding spectral differences (clear/normal and clear/slow modulation indices relative to the conv/normal modulation index) for each talker only in the 1000-Hz octave band. The trends in this octave band are largely representative of all octave-bands, at least for comparisons of clear/slow speech to conv/normal speech: for talkers T3–T5 and to a lesser extent, T1, clear/slow speech has a higher modulation index than conv/normal speech for modulation frequencies up to 3–4 Hz in all octave bands. Such a finding was reported previously (Payton *et al.*, 1994) for clear/slow

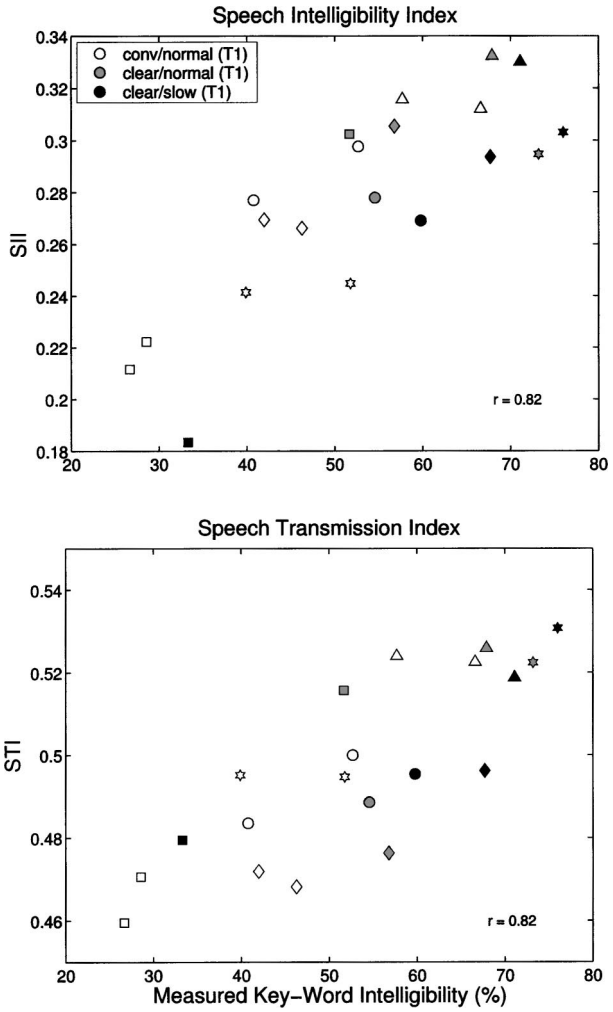


FIG. 4. Top graph shows correlation between the speech intelligibility index (SII) and intelligibility (percent-correct key word scores for normal-hearing listeners in additive speech-shaped noise, SNR=−1.8 dB). Bottom graph shows correlation between the speech transmission index (STI) and measured percent-correct key word intelligibility. Both graphs include data from all five talkers in conv/normal, clear/normal, and clear/slow speaking styles.

speech, but it was unknown whether the enhancement of slowly varying modulations was purely a result of the reduction in speaking rate associated with clear/slow speech. The clear/normal data obtained here suggest that such an increase in modulation does not have to be associated with a change in rate. T3 and T5, for example, show a similar effect on the same range of modulation frequencies (<4 Hz) for clear/normal speech relative to conv/normal speech in a majority of octave-bands (T3: 125-, 500-, 1000-, 2000-, and 4000-Hz bands; T5: 250-, 500-, 1000-, and 2000-Hz bands). Thus, for these two talkers, the intensity envelope spectra of clear/normal speech and clear/slow speech are similar in how they differ from the intensity envelope spectra of conv/normal speech. While the envelope spectra of clear/slow speech for T1 and T4 also differs from the envelope spectra of conv/normal speech in the same way, the envelope spectra of these two talkers' clear/normal speech does not (T4's clear/normal envelope spectra is not substantially different from her conv/normal envelope spectra; T1 exhibits at least as great a modulation increase at frequencies above 4 Hz as below in most octave bands).

Although not every talker exhibited an increase in depth of modulation for low modulation frequencies in clear/normal speech relative to conv/normal speech, recent work by Drullman *et al.*, (1994b,a) suggests that this type of change (observed between clear/slow and conv/normal speech for most talkers and between conv/normal and clear/normal speech for T3 or T5) may contribute to increased intelligibility. These studies demonstrated that modulations as low as 2 Hz are most important for phoneme identification in CVC and VCV syllables. Although a slightly higher range (4–16 Hz) was reported to contribute most to sentence intelligibility, the sentences used in those studies were contextual. Since little context is available when identifying words in nonsense sentences, the syllable data, rather than the sentence data, may be more relevant to the present work.

A strong indication that the increase in lower modulation frequencies may contribute to improved intelligibility stems from the speech transmission index (STI). The STI is a measure of change in modulations of the intensity envelope due to degraded listening conditions, which has been shown to be highly correlated with speech intelligibility (Houtgast and Steeneken, 1985; Humes *et al.*, 1986; Payton *et al.*, 1994). Although the STI is traditionally calculated from the measured change in modulation depth resulting from the transmission of modulated noise signals through an acoustic environment, methods for determining the STI directly from speech waveforms have also been proposed (Houtgast *et al.*, 1980; Houtgast and Steeneken, 1985; Payton and Braida, 1999). One such method was employed to determine the speech-based STI for each talker's conv/normal, clear/normal, and clear/slow speech. In order to compare the results with actual intelligibility scores obtained in Krause and Braida (2002), it was necessary to calculate intensity envelope spectra for each type of speech in the presence of additive noise (SNR=−1.8 dB), as it was presented to listeners in the intelligibility tests, as well as the undegraded speech waveforms. The modulation transfer function (MTF) was then obtained by taking the ratio of the degraded envelope spectra to the undegraded envelope spectra for each of the seven octave bands. From this speech-based MTF, the STI was calculated as in Houtgast and Steeneken (1985). The results of the STI calculations and corresponding intelligibility scores are displayed in Fig. 4. The speech-based STI scores do clearly differentiate the relative intelligibility of talkers and modes. Although the range of STI scores found here (0.46 to 0.53 on a scale of 0 to 1) is compressed relative to the range of intelligibility measures (27% to 76%), the STI is highly correlated with the measured intelligibility scores, with a correlation coefficient of 0.82. Moreover, because the STI score is directly related to envelope spectra, any change in envelope spectra that is associated with an increase in STI should lead to improved intelligibility, at least to the extent that the STI is correlated with intelligibility scores. Therefore, the most salient change in envelope spectra observed here, greater depth of modulation for frequencies <4 Hz, is likely an acoustic property that contributes to the improved intelligibility of clear/normal speech, given that the STI of

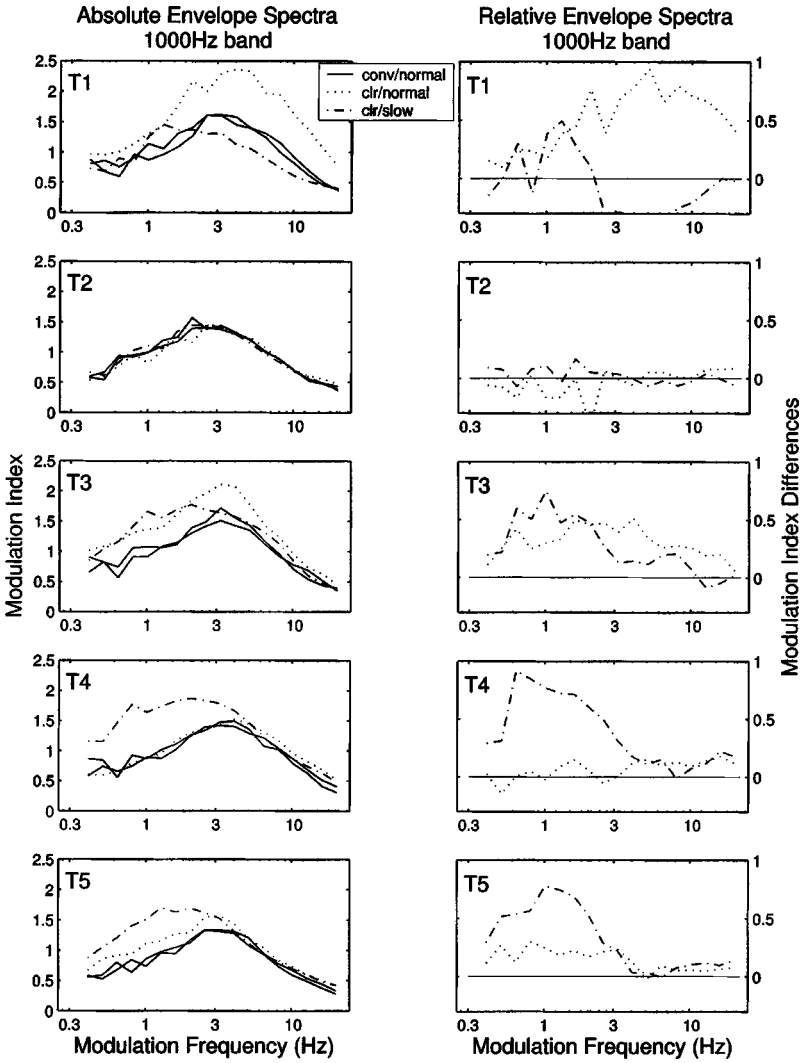


FIG. 5. Temporal envelope modulation data. Graphs at left show absolute intensity envelope spectra for each talker and speaking style, depicted by modulation index (indicating depth of modulation) as a function of third-octave band modulation frequency. This measurement was made for seven octave bands; the 1000-Hz octave band is shown here. Corresponding graphs at right show envelope spectral differences, obtained by subtracting the conv/normal modulation spectrum from the clear/normal and clear/slow modulation spectra, which depicts the relative distribution of modulation spectral energy between conversational and clear speech.

T5 (a talker who exhibited this change) increased from 0.49 for conv/normal speech to 0.52 for clear/normal speech and 0.53 for clear/slow speech.

B. Phonological measurements

In running speech, speech segments in certain contexts undergo specified transformations described by phonological rules. In order to characterize how frequently such phonological phenomena occur in clear and conversational speech, it was first necessary to label utterances from the database, both phonetically and phonemically. Phonetic labels were determined from a combination of information obtained through aural repetitions and spectrograms. These phonetic labels were then aligned with the corresponding dictionary pronunciation of the sentence, so that phonemes could be compared across speaking mode and talker.

As discussed previously, phonological measurements were restricted to two talkers, T4 and T5 (see Table I). The frequency of phonological phenomena occurring in clear/normal, clear/slow, and conv/normal speech was measured for these two talkers as a percent of total possible opportunities for such phenomena. Similar to Picheny *et al.* (1986), the phenomena were classified into the following categories:

(1) **Vowel modification (VM)**: This category refers to vowel substitutions and the reduction of vowels to a schwa in unstressed syllables. Also included in this category is the replacement of a vowel-sonorant sequence with a syllabic version of the sonorant. Total possible opportunities were defined as the total count of vowels in function words, unstressed vowels in content words, and vowel-sonorant sequences. Of the four lists that were examined, the number of possible opportunities ranged from 187 to 197.

(2) **Burst elimination (BE)**: Burst elimination is the failure to create a burst at the release of a stop consonant. It is particularly common when the stop consonant is in word-final position or syllable-final position. Total possible opportunities were designated as the number of sentence-final stop consonants and word-final stop consonants followed by another consonant. The number of opportunities ranged from 66 to 79 across lists.

(3) **Alveolar flap (AF)**: This phenomenon occurs most frequently when a /t/ or /d/ is preceded by a stressed vowel and followed by an unstressed vowel. The /t/ or /d/ is frequently realized as a flap (e.g., ''latter,'' ''ladder''). Total possible opportunities consisted of all instances of /d/ or /t/ occurring between two vowels. The lists that

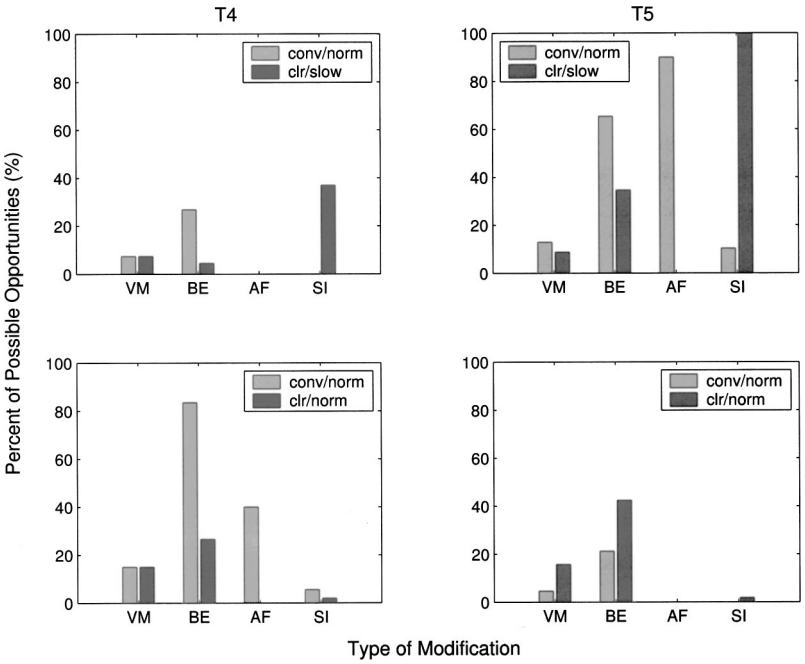


FIG. 6. Frequency of phonological phenomena (**VM**=vowel modification, **BE**=burst elimination, **AF**=alveolar flap, **SI**=sound insertion). Top row compares sentences spoken in both conv/normal and clear/slow modes, and bottom row compares sentences spoken in both conv/normal and clear/normal modes. Columns give results for each talker. Each row is a different corpus of sentences. Frequency for each category was calculated as a percent of possible opportunities. For each corpus of sentences, the number of possible opportunities was as follows: VM=187–197, BE=66–79, AF=5–10, and SI=201–233.

were examined had relatively few opportunities for flaps, and the number of possible opportunities ranged from five to ten.

(4) **Sound insertion (SI)**: This category refers to the insertion of a schwa vowel after a voiced consonant. The insertions frequently occur at the end of words, particularly in clear/slow speech (Picheny *et al.*, 1986). Total possible opportunities were set to be the total number of syllables in the list, and the number of opportunities ranged from 201 to 233 across lists.

Figure 6 shows percentage histograms for these types of phonological modifications for both talkers in all of the speaking styles. Every histogram compares the same sentence list spoken in two different styles, but each of the four histograms is measured on a different list. Therefore, comparisons across sentence lists (i.e., across different histograms) is inappropriate. Note that markedly different scores are obtained for each talker's conv/normal speech on the two different sentence lists. In the case of alveolar flap scores, such variation for conv/normal speech is likely a result of the small number of possible opportunities (ranging from five to ten for the various lists); however, the variation in other modifications within speaking mode (across sentence list) cannot be accounted for with this reasoning. The most likely explanation is that global factors such as word position within the sentence influences a talker's decision to eliminate the burst. In addition, it is likely that the method for counting total possible opportunities was somewhat crude. Despite these limitations, comparisons of modifications between modes on the same sentence list (i.e., within histogram) should nonetheless have value, so we concentrate on those comparisons here.

Comparing within sentence list, the results for T5's clear/slow speech are similar to those reported by Picheny *et al.* (1986): the frequency of vowel modifications, eliminations of stop bursts, and alveolar flaps was smaller in clear/slow speech than in conv/normal speech, and the relative number of sound insertions was much larger. The results for T4's clear/slow speech, however, show these trends only for elimination of stop bursts and sound insertions; her percentage of vowel modifications and alveolar flaps were unchanged between conv/normal and clear/slow modes. It is interesting to note that while both talkers showed a large increase in sound insertions in clear/slow speech relative to conv/normal speech, their clear/normal speech had about the same number of sound insertions as conv/normal speech. Thus, it would appear that the sound insertions found in clear/slow speech are not essential to improvements in intelligibility (relative to conv/normal speech).

Perhaps the most striking finding regarding phonological modifications is that, while the trend for T4 involving vowel modification, burst elimination, and alveolar flaps in clear/normal speech is similar to that typically found in clear/slow speech, T5's data actually exhibits the opposite effect. The percentage of vowel modifications and burst eliminations in T5's clear/normal speech is higher than in his conv/normal speech. One possible explanation for this result is that the occurrences of these phenomena may have been concentrated primarily in function words, which were not scored in the intelligibility tests.

C. Phonetic measurements

Phonetic level measurements included segmental power, segmental phone duration, short-term rms spectra, vowel formant frequencies, consonant–vowel ratio, and voice-onset time. Formant transition duration and extent was also examined, since some recent work (Ochs *et al.*, 1989; Tallal *et al.*, 1996; Tallal and Piercy, 1975) shows this parameter to be an important determinant of speech intelligibility. Unless otherwise stated, phonetic measurements were made for each phonetic segment listed in Table III, and segment boundaries

TABLE III. Phonetic segments included in phonetic analysis. Vowels and consonants are represented with the appropriate IPA symbol; other phonetic segments are represented with abbreviations indicated below.

Category	Segments	
Vowels	ɑ, æ, ʌ, ɔ, aʊ, ə, aɪ, ɛ, ɝ, eɪ, ɪ, i, oʊ, ɔɪ, ʊ, u, l̩, m̩, n̩	
Consonants	b, d, g, p, t, k, ɾ, w, ʍ, j, l, ɹ, m, n, ŋ, f, θ, s, ʃ, h, v, ð, z, ʒ, tʃ, dʒ	
Other	sil	silence
	bst	stop burst
	cl	voiceless closure
	vcl	voiced closure

were derived from the label files. Again, these measurements were restricted to two talkers, T4 and T5 (see Table I).

1. Power

To compare the relative power of phones spoken in conv/normal and clear/normal modes, the rms level for each phonetic segment in both modes was computed. Statistical comparisons of the conv/normal and clear/normal levels for each phone were performed via paired *t*-tests, where the difference variable $D=L_{clear}-L_{conv}$ was formed from L_{conv}, the level for the phone spoken conversationally, and L_{clear}, the level for the phone spoken clearly.

Few phones for either talker (12 for T4, 8 for T5) showed a statistically significant difference ($p=0.05$) in level, considering that rms levels were measured for 49 phones. Moreover, only three phones (/m/, /n/, and /w/) changed significantly for both talkers. Furthermore, most changes in level were small, often less than 1 dB and only greater than 3 dB in four cases (T4's /m/: $D_{avg}=3.6$; T4's /n/: $D_{avg}=4.0$; T4's /θ/: $D_{avg}=-5.9$; T4's /v/: $D_{avg}=4.7$. This result is not particularly surprising, since the long-term rms level of each sentence was normalized. For a talker to produce a particular phone with greater power in clear/normal speech relative to conv/normal speech would mean that one or more of the other phones would have to be produced with relatively less power. Talkers may have found experimentally that modifying relative segmental power was not beneficial to intelligibility, or perhaps the mental load and articulatory demands of producing speech in this manner was prohibitively difficult. In either case, it appears that a change in relative power of phonetic segments is not integral to highly intelligible speech.

The average energy of phones in conv/normal and clear/normal speech was also calculated from pressure and duration measurements, according to the formula $E=P^2\times T$, where E is energy, P is rms pressure of the phonetic segment, and T is the length of the phonetic segment in seconds. The relative energy in each speaking style was compared for each phone via paired *t*-tests, where the difference variable was defined as $D=E_{clear}-E_{conv}$, the difference between the energy of the phone spoken conversationally (E_{conv}) and the energy of the phone spoken clearly (E_{clear}). As with power, few phones for either talker (10 for T4, 8 for T5) showed a statistically significant difference ($p=0.5$) in energy. Of phones that differed significantly, slightly more than half for each talker showed an increase in energy. However, the increase was often small (less than 3 dB), and only four phones changed significantly for both talkers (/m/: $D_{avg}=6.4$ for T4, $D_{avg}=5.3$ for T5; /ɹ/: $D_{avg}=2.4$ for T4, $D_{avg}=1.4$ for T5; /ʊ/: $D_{avg}=6.0$ for T4, $D_{avg}=-3.5$ for T5; and /w/: $D_{avg}=6.9$ for T4, $D_{avg}=5.7$ for T5).

2. Duration

The relative durations of phones spoken in conv/normal and clear/normal modes were also computed and compared statistically via paired *t*-tests. In this case, the difference variable was $D=D_{clear}-D_{conv}$, where D_{conv} was the duration of the phone spoken conversationally, and D_{clear} was the duration of the phone spoken clearly.

Again, not a great many phones of the 49 measured showed a statistically significant difference ($p=0.05$) in duration for either talker (10 for T4, 15 for T5). Moreover, although the phones that did change duration significantly for both talkers were primarily vowels, there was not a consistent pattern of change across talkers. Only three phones (/ʌ/, /ɛ/, and /z/) exhibited significant changes for both talkers; moreover, the changes in duration for /ʌ/ and /ɛ/ were in opposite directions for each of the talkers. This trend was true for most phones and for all vowels with a significant change in duration: T4's eight statistically significant vowels were shortened in clear/normal speech relative to conv/normal speech, while T5's seven statistically significant vowels were lengthened.

The above measurements, however, were made for all phones, independent of phonetic context. Since it is well known that vowels preceding voiceless consonants (in the same syllable) are typically shorter than vowels preceding voiced consonants (Ladefoged, 1975; Denes, 1955), durational measurements were also made taking context into account. Durations were measured for each vowel and for all vowels as a group, when followed by either voiced phones (/z/, /ʒ/, /v/, /ð/, /b/, /d/, and /g/) or unvoiced phones (/s/, /ʃ/, /f/, /θ/, /p/, /t/, and /k/). Durations were compared statistically with paired *t*-tests as above. The results were quite similar to the context-independent duration measurements, with T4's statistically significant changes in clear/normal speech generally showing a decrease in duration and T5 showing an increase.

3. Short-term rms spectra

The short-term spectra of conv/normal and clear/normal speech, normalized for segment rms level, were computed. For both normal rate speaking modes, spectra corresponding to each phone were averaged over every occurrence of the phone in 50 sentences. The spectra were computed by averaging FFTs of windowed portions of the signal using 25.6-ms Hamming windows, incremented in 1-ms intervals. The spectra were also passed through a preemphasis filter with a slope of +6 dB/octave in order to boost the higher frequencies. Finally, a $\frac{1}{3}$-octave representation of the spectra was obtained by summing components over $\frac{1}{3}$-octave intervals with center frequencies ranging from 62.5 to 8000 Hz.

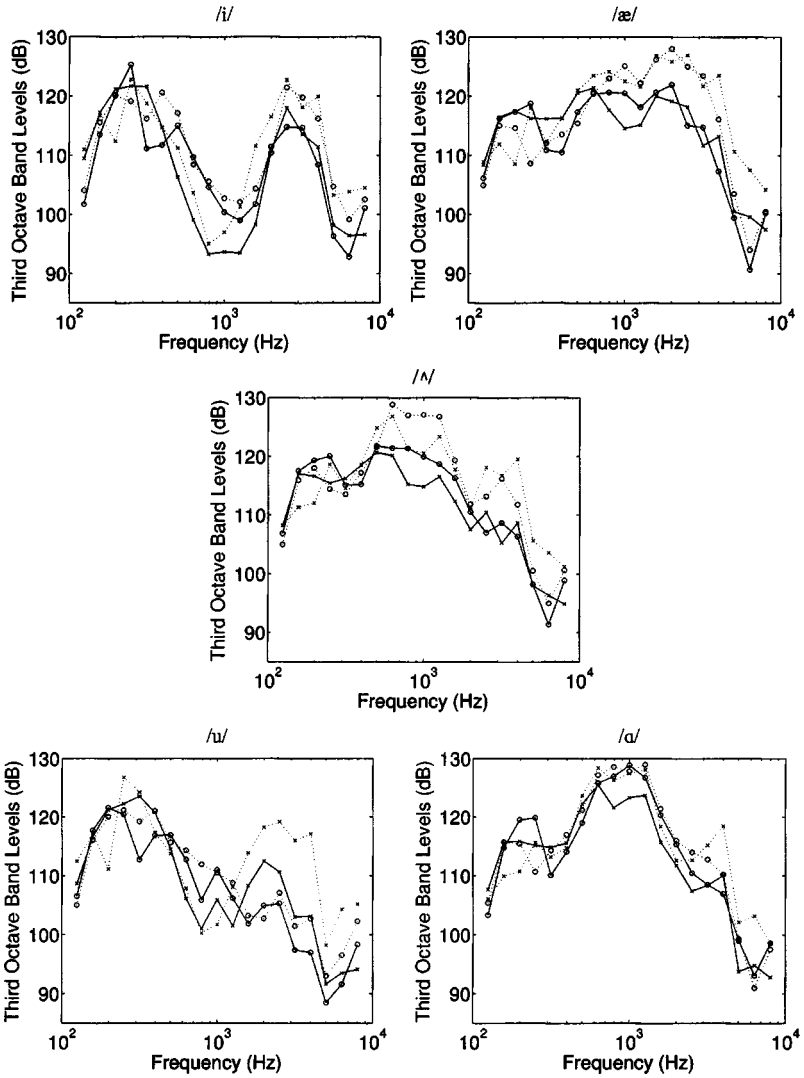


FIG. 7. Third-octave average spectra for five vowels (spanning F1–F2 space) in conv/normal and clear/normal modes for both talkers. Solid lines are conv/normal data, and dotted lines are clear/normal data; circles (○) represent T4, and exes (x) represent T5. Similar results were obtained for most vowels.

Although neither talker exhibits much spectral change between conv/normal and clear/normal modes for consonants, qualitative changes in vowel spectra are apparent. Typical average vowel spectra for each talker in both modes are shown in Fig. 7. The vowels displayed span the range of F1 (correlated with vowel height) and F2 (correlated with vowel fronting) and thus are fairly representative of the entire vowel space. These plots show that both talkers have higher spectral prominences at the vowel formant frequencies in clear/normal speech than in conv/normal speech, which has also been confirmed by inspection of individual vowel tokens. The result holds for nearly all other vowels as well. The effect appears strongest for the second and third formants and is somewhat stronger for T5 than for T4. Such a spectral emphasis of formants is likely responsible for the observed long-term spectral differences, since vowels are relatively long in duration compared to consonants and thus contribute substantially to the long-term spectrum. Moreover, it is possible that this emphasis led to enhanced intelligibility of vowels in clear/normal speech, since vowel formants have long been considered one of the most salient cues to vowel identification (Ladefoged, 1975; Peterson and Barney, 1952; Hillenbrand *et al.*, 1995). However, the correlation between intelligibility of individual vowel tokens and their corresponding short-term spectra would be necessary to verify this possibility, and such an undertaking is too labor-intensive for the purposes of this initial acoustic analysis.

4. Vowel formant frequencies and formant bandwidths

Vowel formant frequencies and formant bandwidths were measured using the formant tracking program provided in the *ESPS/waves+* software package. The first three formants and their bandwidths were extracted at the midpoint of each vowel and averaged over all occurrences of the vowel within a speaking style. Paired *t*-tests were employed to compare statistically the formants and formant bandwidths in conv/normal and clear/normal speaking modes. The difference variable for formant bandwidth comparisons was $D = BW_{clear} - BW_{conv}$, and the difference variable for formant frequency comparisons was $D = F_{clear} - F_{conv}$, where F represents formant frequency, BW represents formant bandwidth, and subscripts indicate the speaking mode.

Of the few bandwidths that changed significantly ($p = 0.05$) for either talker (4 for T4, 12 for T5), all were slightly narrower in clear/normal speech than in conv/normal speech. One potential benefit of narrower bandwidths, par-

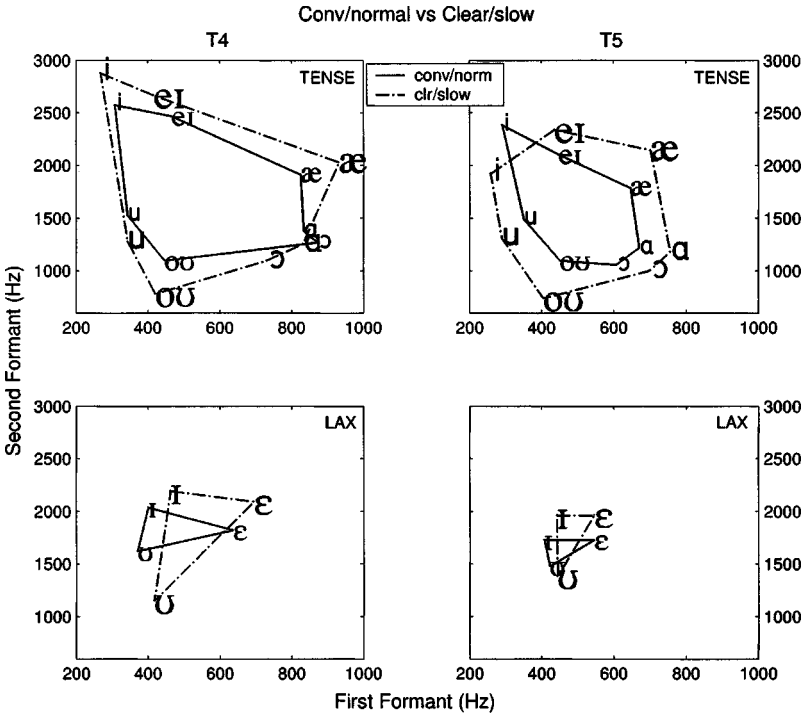


FIG. 8. Vowel formant frequency data. Top row shows results for tense vowels and bottom row for lax vowels; columns give results for each talker. Conv/normal speech is indicated by smaller phonetic symbols and solid lines, and clear/slow speech is indicated by larger phonetic symbols labels and dash-dotted lines. Note that the vowel space is expanded in clear/slow speech relative to conv/normal speech.

ticularly in conjunction with the higher formant amplitudes found in the short-term spectra, is that more energy is concentrated very near the vowel formant frequencies, which may aid listeners in identifying vowels. Very few (2–3) second or third formants exhibited a statistically significant ($p = 0.05$) frequency change from conv/normal to clear/normal speech for either talker, and none of these were significant for both talkers. Slightly more (3–6) first formants exhibited a statistically significant frequency change between modes for each talker. In the majority of these cases where F1 differed significantly between modes, a higher frequency was observed for clear/normal speech than for conv/normal speech. This result is consistent with the fact that clear/normal speech was produced at higher intensities than conv/normal speech, since louder speech is typically obtained with a larger jaw opening, resulting in decreased tongue height and increased F1 (Huber *et al.*, 1999).

Plots and subsequent discussion are limited here to the first two formants, in order to facilitate comparison of changes in formant frequencies with previous work and because results for the third formant were similar. Figure 8 demonstrates that, as in previous studies (Picheny *et al.*, 1986; Chen, 1980), the vowel space for clear/slow speech is expanded relative to conv/normal speech. The expanded vowel space relative to conv/normal speech, however, is less evident for clear/normal speech in Fig. 9. It can be seen only for tense vowels produced by T5.

Since substantial changes in formant frequencies were not observed, one hypothesis was that there could be a tighter clustering of formants in clear/normal speech relative to conv/normal speech, a phenomenon observed by Chen (1980) in clear/slow speech. To test this hypothesis, a *t*-test was employed, using the test statistic for variances associated with paired observations (Glass and Hopkins, 1968):

$$t = \frac{s_1^2 - s_2^2}{2 s_1 s_2 \sqrt{(1 - r_{12}^2)/v}},$$

where s_1^2 and s_2^2 are the sample variances of conv/normal and clear/normal formants, respectively, r_{12} is the correlation coefficient between paired observations, and v represents the degrees of freedom.

Slightly less than one-fourth (10 for T4, 10 for T5) of the 48 formant frequencies (16 vowels×3 formants) for each talker exhibited a statistically significant change in variance. Of these, more than half (7 out of 10 for each talker) showed a variance increase from conv/normal to clear/normal modes. Thus, there is not a trend in the data indicating that formant frequencies in clear speech cluster more tightly than in conversational speech, as was reported previously for clear/slow speech by Chen (1980). The differences between Chen's data and this study are most likely a result of Chen's use of CVC materials, which introduces less contextual variability into the vowels than sentence materials. The data in this study are similar to the results for clear/slow speech that were reported by Picheny *et al.* (1986), who also used sentence materials. Formant frequency clustering was also examined for vowels in key words and in word-initial positions and similar results were obtained.

In summary, vowel formants extracted at the midpoints of vowels did not appear to cluster more tightly or to be significantly closer to targets in clear/normal speech than in conv/normal speech. While it was hypothesized that formant frequencies would change in a way that would emphasize or exaggerate articulatory targets in clear/normal speech, it cannot generally be seen in the data. One possible explanation for this deviation from the theoretical prediction is that the formants were measured only once, at the vowel midpoint. There is growing evidence that listeners rely on formant fre-

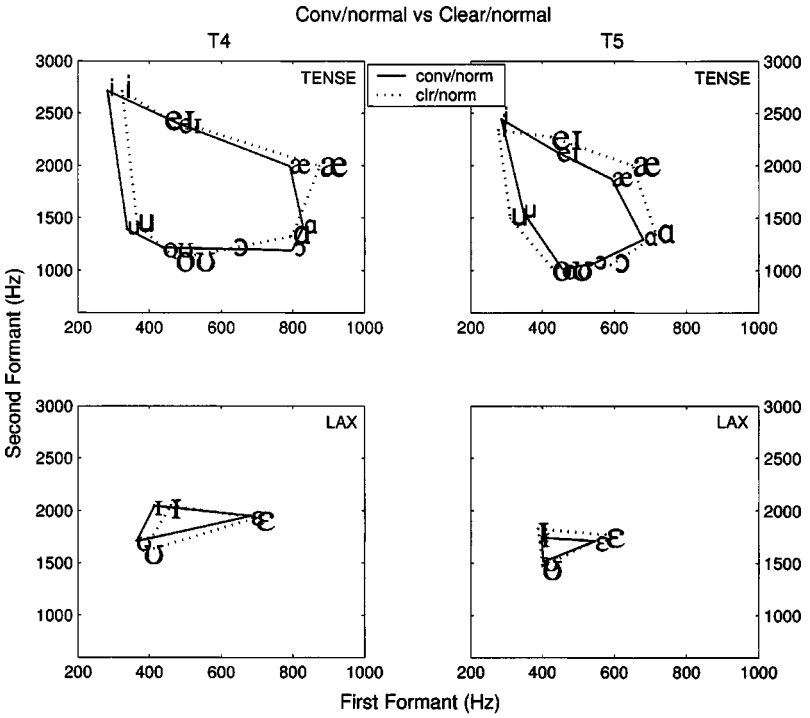


FIG. 9. Vowel formant frequency data. Top row shows results for tense vowels and bottom row for lax vowels; columns give results for each talker. Conv/normal speech is indicated by smaller phonetic symbols and solid lines, and clear/normal speech is indicated by larger phonetic symbols labels and dash-dotted lines. An expanded vowel space for clear/normal speech relative to conv/normal speech is not apparent.

quency movements throughout the vowel and not just at CV and VC transitions (Hillenbrand and Nearey, 1999). It is possible that formant contours in clear/normal speech more closely matched target values, but cannot be sufficiently described by a measurement of the formant frequencies at one point in time. In any case, changes in steady-state vowel formant locations are not likely to account for the high intelligibility of clear/normal speech.

5. Formant transitions: Duration and extent

Because some studies (Ochs *et al.*, 1989; Tallal *et al.*, 1996; Tallal and Piercy, 1975; Turner *et al.*, 1997) indicate that formant transition duration and extent may play an important role in speech intelligibility, an attempt was also made to collect formant transition data (duration, rate, beginning and ending frequency values, and extent of formant transitions) from conv/normal and clear/normal speech. CV sequences beginning with stop consonants (/p/, /t/, /k/, /b/, /d/, and /g/) were examined, but with only about 100 CV tokens beginning with stop consonants in the database, there were not many instances of each CV pair to be analyzed. Even when consonants and vowels were grouped according to place of articulation, there were very few data points to be analyzed for each pair of C and V groups, and the results were inconclusive.

6. Consonant–vowel ratio

Much attention in the literature (Montgomery and Edge, 1988; Gordon-Salant, 1987) has been given to consonant-vowel ratio (CVR), or ratio of consonant power to the power in the nearest adjacent vowel. Picheny *et al.* (1986) did not measure consonant-vowel ratio *per se* and criticized the utility of CVR measurements in sentential environments. That study did, however, report an increase in the relative power of some, particularly unvoiced, consonants. Since such an increase in power was not observed here for clear/normal speech, CVR was also measured for the sake of completeness.

Consonants were grouped into five classes: stops, fricatives, affricates, liquids, and nasals. For stop consonants, the power was designated as the power in the stop burst. Affricate power was defined as the nonclosure power. For all other consonant classes and for vowels, the rms level was used. Resulting CVRs reveal little differences between speaking styles for any of the consonant classes. The only possible exception is affricates, which have a 2–3 dB larger CVR in clear speech than in conversational speech for both talkers.

7. Voice-onset time

Voice-onset time (VOT) is widely considered an important cue for voicing contrasts in stop consonant identification (Ladefoged, 1975; Edwards, 1981). Therefore, VOTs were measured for word-initial stop consonants in conv/normal and clear/normal speech. Due to the nature of the sentence database, such word-initial stop consonants appeared only in content words. The VOT data was tabulated through a semi-automated procedure: using the *ESPS/waves+* software package, a human labeller marked the release of the stop burst and the onset of voicing, relying on information available in the time waveform and the spectrogram to judge when voicing had begun. The VOT was then automatically calculated and recorded. Statistical comparisons of the conv/normal and clear/normal VOTs for each stop consonant were then performed via paired *t*-tests, where the difference variable $D = VOT_{clear} - VOT_{conv}$ was formed from VOT_{conv} and VOT_{clear}, the conv/normal and clear/normal VOTs, respectively.

Mean VOTs for each stop consonant in conv/normal and clear/normal speech and corresponding difference variables

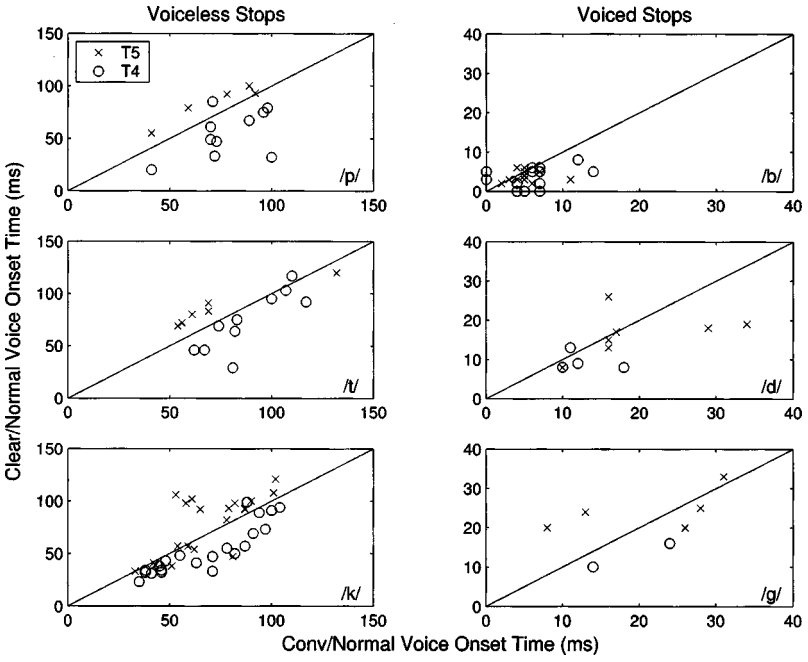


FIG. 10. Comparison of stop consonant voice-onset time in clear and conversational modes at normal speaking rates.

that were statistically significant ($p=0.05$) are listed in Table IV. In addition, a comparison of each token's clear/normal and conv/normal VOTs is shown in Fig. 10. The overall results are in good agreement with average VOTs previously reported for stop consonants, showing clearly that unvoiced stops have longer VOTs and that VOT increases as place of articulation moves from labial to alveolar to velar. Statistically significant changes in VOT between conv/normal and clear/normal speech appear only in voiceless stop consonants. The talkers once again exhibit opposite trends: T5's data reveals an increase in VOT in clear/normal speech, while T4's data shows a decrease. Although VOTs have not been previously reported for clear/slow speech, the expected trend would be similar to that of T5. First, slower articulation rates such as that used for clear/slow speech have been shown to increase VOTs for voiceless stop consonants while leaving VOTs of voiced stop consonants unchanged (Summerfield, 1981). Second, such a change in VOTs for voiceless consonants would presumably increase the perceptual distance between each pair of consonants that differs only by the voicing distinction.

III. ASSESSMENT OF ACOUSTICAL RESULTS

Since clear/normal speech exhibits a sizable (14 point) intelligibility advantage over conversational speech, it can be assumed that at least some of the essential acoustical characteristics of clear speech are preserved in clear/normal speech. However, since the differences obtained in the acoustical analysis were not dramatic and were fewer in number than the differences obtained between conv/normal and clear/slow speech, there was some concern that these differences may not be able to account for a large part of the intelligibility advantage found in clear/normal speech. One explanation of the observed acoustic characteristics is that the essential characteristic of clear speech may be a linear combination of several acoustical parameters. In this case, it is possible that comparing individual acoustic characteristics, as in the previous section of this paper, would prove inadequate for capturing the difference between speaking modes, which would account for the fact that fewer individual acoustic differences were found between conv/normal and clear/normal speech.

In order to assess whether a multivariate linear combination of the acoustic variables could account for the difference between conv/normal and clear/normal speech, a discriminant analysis was performed on the acoustic data. The analysis was limited to T4 and T5, the two talkers for whom all measures (global, phonological, and phonetic) were available. For both talkers, the analysis was performed on data for classes of phonemes, with the vowel class being broken into subclasses based on place of articulation: HighFront (/ɪ/ and /i/), MidFront (/ɛ/ and /eɪ/), LowFront (/æ/), Neutral (/ʌ/ and /ə/), HighBack (/ʊ/ and /u/), and LowBack (/ɑ/, /ɔ/, and /oʊ/). For classes containing vowels or voiced phonemes, the discriminating variables were duration, power, F0, the first four

TABLE IV. Mean values of VOT for conv/normal and clear/normal speech. Diff column shows only those differences that were significant in paired *t*-tests at the $p=0.05$ level.

Talker		Conv/norm (ms)	Clr/norm (ms)	Diff (ms)
T4	/p/	78	55	−23
	/t/	88	74	−15
	/k/	68	53	−15
	/b/	5.6	3.6	⋯
	/d/	13	9.5	⋯
	/g/	19	13	⋯
T5	/p/	72	84	+12
	/t/	74	86	+12[a]
	/k/	68	76	+9.2[b]
	/b/	5.1	3.9	⋯
	/d/	20	17	⋯
	/g/	21	24	⋯

[a]Significance level: $p=0.057$.
[b]Significance level: $p=0.055$.

TABLE V. Summary of acoustical differences: clear/normal speech relative to conv/normal speech.

Property	Difference	Talker
Global properties (measured for all talkers):		
F0	greater average, range[a]	T5
Long-term spectra	more energy above 1 kHz	T2, T3, T4, T5
Temporal envelopes	larger modulation index for frequencies <3–4 Hz[b]	T3, T5
Phonological properties (measured for T4, T5 only):		
Word-final stops	bursts released more often	T4
	bursts released less often[c]	T5
Phonetic properties (measured for T4, T5 only):		
Short-term vowel spectra	increased energy near second and third formants	T4, T5
Word-initial voiceless stops	decreased VOT[c]	T4
	increased VOT	T5

[a]Difference was present only for male talker.
[b]Difference was achieved in 5 of 7 octave bands by T3 and 4 of 7 octave bands by T5.
[c]Difference was in opposite direction of that typically expected for clear/slow speech.

formants, and their corresponding bandwidths. For classes containing voiceless phonemes, the discriminating variables were limited to only duration and power.

Unfortunately, the results of the discriminant analysis did not lend additional insight into the data. The technique was largely unsuccessful in discriminating T4's conv/normal and clear/normal tokens, as evidenced by the fact that the smallest Wilk's lambda (a measure of the variance not explained by the discriminating function) obtained for any phoneme class was 0.45. While the analysis did discriminate fairly well between most classes of T5's conv/normal and clear/normal vowels, the most heavily weighted discriminating variable was consistently F0. Moreover, the weightings of the other variables across vowel class were quite variable. Therefore, the results of the discriminant analysis for T5 merely confirmed one of the global phenomena reported in the previous section of this paper: T5 exhibits an increase in F0 average in clear/normal speech relative to conv/normal speech.

IV. DISCUSSION

Table V lists the measured acoustic differences between clear/normal and conv/normal speech. It is evident that the acoustic properties of clear/normal speech for T4 and T5, the two talkers whose speech was analyzed at all three levels of detail (global, phonological, and phonetic), differed dramatically (e.g., VOT changed in opposite directions), suggesting that talkers may employ different strategies for producing clear speech at normal rates. However, one similarity between the talkers was also observed: the talkers omitted a number of acoustic properties in clear/normal speech that were present in clear/slow speech. The properties that were omitted, however, varied for each talker. T4 omitted expected changes in temporal envelopes, F0, and VOT, while T5 omitted expected changes in stop burst release frequency (both talkers also omitted a number of other differences commonly found between conv/normal and clear/slow speech; most notably, more frequent and longer pauses and vowel formant frequencies which more closely approximate target values). Thus, it appears that while most talkers (here and in previous clear speech studies) use the same or similar strategies when producing clear/slow speech, many of them may use different strategies when producing clear/normal speech. The omissions in clear/normal speech are most likely caused by physiological constraints on articulation at normal speaking rates that hinder simultaneous expression of such a large number of acoustic properties as typically found in clear/slow speech. However, the reason for the variation in types of omissions across talker is not known. Evaluations of signal processing schemes based on acoustic characteristics of various talker strategies for clear/normal speech should lend additional insight into this question.

If different talker strategies for producing clear/normal speech do exist, it would then be reasonable to expect that additional types of clear speech could be elicited from talkers by varying either the type of distortion employed or the constraints imposed on speaking rate. Furthermore, the intelligibility of each type of clear speech may then vary for different distortions. A similar effect has been shown for different styles of conversational speech: one style of conversational speech can be significantly more intelligible than another style in reverberation, even though the two styles of speech have comparable intelligibility under low-pass filtering (Dix, 1996). The question of whether and how many types of clear speech exist has yet to be explored. This work is not intended to characterize all types of clear speech but, rather, to focus on determining the properties of the specific type(s) of clear/normal speech that have been produced in Krause and Braida (2002) and shown to provide an intelligibility advantage to normal hearing listeners with simulated hearing losses, achieved by additive noise.

For the types of clear/normal speech analyzed in this study (resistant to additive noise), a number of acoustical differences between conv/normal and clear/normal speech were identified (see Table V). Yet, it is important to keep in mind that (1) these acoustic characteristics are not likely to contribute equally to the intelligibility advantage of clear/normal speech, (2) it is possible that some of these charac-

teristics do not contribute at all to improved intelligibility, and (3) there may be additional characteristics which do contribute to improved intelligibility but were not identified with the measurements made in this study. One factor that may have prevented identification of one or more additional properties of clear/normal speech is the structure of the sentences in the acoustic database. The sentence database was well-suited for the task of eliciting clear/normal speech from talkers in the initial intelligibility study (Krause and Braida, 2002), providing talkers with a natural, familiar task and the opportunity to manipulate the acoustics of an entire sentence instinctively, including global and phonological variables as well as phonetic variables. For the purposes of acoustic analysis, however, the database poses at least two potential difficulties.

The first difficulty in performing acoustic measurements on a database of nonsense sentences is that acoustic properties that are typical of more complex, meaningful utterances are not likely to be present in such a database. Yet, identifying changes in these properties may well be important for improving intelligibility in realistic situations. On the other hand, the second difficulty is that the complexity of the sentence database leads to the presence of acoustic variability, due to word positioning within sentences or phonetic context within words. It seems possible that for some acoustic properties, this variability could be large enough to mask the variability between conv/normal and clear/normal tokens. For example, in one sentence list, the 25 occurrences of /ɛ/ in content words appeared in 22 unique phonetic environments, if the phonetic context of both the preceding and subsequent phone is considered. Thus, when the tokens were analyzed as a group, token variability due to environment may have been large enough to mask variability due to speaking mode, even when content words were analyzed separately from function words. However, with so few tokens in each phonetic environment, a meaningful statistical analysis of speaking mode differences within specific environments could not be performed. Therefore, a similar acoustic analysis on a database with only a fixed set of environments and many tokens in each environment may lead to identification of additional phonetic-level properties associated with clear/normal speech in addition to those identified in this study.

Despite the limitations of the database, the differences summarized in Table V could each potentially be characteristics that improve speech intelligibility. In particular, a strong case can be made that two of the global properties of clear/normal speech identified in this study are linked to significant intelligibility gains. First, changes in long-term spectra were exhibited by four of five talkers in this study, and the SII, which was highly correlated with intelligibility scores measured in Krause and Braida (2002) indicates that such changes are likely to lead to increased intelligibility. Second, the STI, which was also highly correlated with measured intelligibility scores, indicates that the changes in envelope spectra exhibited by T5 should lead to increased intelligibility. Thus, these properties of clear/normal speech seem reasonably likely to contribute to its high intelligibility.

The remaining phonological and phonetic differences listed in Table V include phonological properties that changed by 15 percentage points or more and phonetic properties that showed a statistically significant difference between conv/normal and clear/normal modes in paired *t*-tests. In addition, the difference in short-term vowel spectra is included because it seems likely to be related to the difference in long-term spectra. While each of these properties has the potential to contribute to increases in intelligibility, further investigation is necessary in order to identify those properties most closely associated with improvements in intelligibility. For example, analysis of correlations between intelligibility of individual phonemes in the database and corresponding phonetic characteristics would be helpful in this regard.

Lastly, it should be noted that a different set of acoustic properties associated with clear/normal speech might have been identified if different talkers had been selected for detailed (phonological and phonetic level) analysis or if different sentence materials had been used when the speech was elicited. Given that T4 and T5 use differing strategies, it is reasonable to assume that other talkers may also employ unique strategies for producing clear/normal speech. While additional insights into the properties of clear/normal speech may subsequently be obtained by acoustic analysis of additional talkers, the analysis of the two talkers described here provides an initial pool of acoustic properties to be investigated further for their contributions to highly intelligible speech. One possible method for investigating the extent to which each of these acoustic properties contribute to improved intelligibility is to conduct intelligibility tests that evaluate these characteristics singly and in combination. For constructing such intelligibility tests, signal processing transformations that manipulate only these characteristics would be useful. The information gained from such intelligibility tests could ultimately lead to better signal processing approaches for digital hearing aids.

ACKNOWLEDGMENTS

The authors wish to thank Jae S. Lim, Joseph S. Perkell, Kenneth N. Stevens, and Rosalie M. Uchanski for many helpful technical discussions. In addition, we thank M. Keller and D. Rhoads for their assistance with phonetic labelling of the speech materials. Financial support for this work was provided by a grant from the National Institute on Deafness and Other Communication Disorders (NIH Grant No. 5 R01 DC 00117) and a National Defense Science and Engineering Graduate fellowship from the Office of Naval Research.

ANSI-S3.5-1997 (**1997**). ''American national standard methods for the calculation of the speech intelligibility index,'' J. Acoust. Soc. Am. 10.

Bradlow, A. R., Krause, N., and Hayes, E. (**2003**). ''Speaking clearly for children with learning disabilities: Sentence perception in noise,'' J. Speech Hear. Res. **46**, 80–97.

Chen, F. R. (**1980**). ''Acoustic characteristics and intelligibility of clear and conversational speech,'' Master's project, MIT, Cambridge, MA.

Cutler, A., and Butterfield, S. (**1990**). ''Durational cues to word boundaries in clear speech,'' Speech Commun. **9**, 485–495.

Cutler, A., and Butterfield, S. (**1991**). ''Word boundary cues in clear speech: A supplementary report,'' Speech Commun. **10**, 335–353.

Denes, P. (**1955**). ''Effect of duration on perception of voicing,'' J. Acoust. Soc. Am. **27**, 761–764.

Dix, A. K. (**1996**). Personal communication.
Drullman, R., Festen, J. M., and Plomp, R. (**1994a**). ''Effect of reducing slow temporal modulations on speech reception,'' J. Acoust. Soc. Am. **95**, 2670–2680.
Drullman, R., Festen, J. M., and Plomp, R. (**1994b**). ''Effect of temporal envelope smearing on speech reception,'' J. Acoust. Soc. Am. **95**, 1053–1064.
Edwards, T. J. (**1981**). ''Multiple features analysis of intervocalic english plosives,'' J. Acoust. Soc. Am. **69**, 535–547.
Glass, G. V., and Hopkins, K. D. (**1968**). *Statistical Methods in Education and Psychology* (Academic, New York).
Gordon-Salant, S. (**1987**). ''Effects of acoustic modification on consonant recognition by elderly hearing-impaired subjects,'' J. Acoust. Soc. Am. **81**, 1199–1202.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of american english vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.
Hillenbrand, J. M., and Nearey, T. M. (**1999**). ''Identification of resynthesized /hvd/ utterances: Effects of formant contour,'' J. Acoust. Soc. Am. **105**, 3509–3523.
Houtgast, T., Steeneken, H., and Plomp, R. (**1980**). ''Predicting speech intelligibility in rooms from the modulation transfer function,'' Acustica **46**, 60–72.
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.
Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., and Johnson, K. (**1999**). ''Formants of children, women, and men: The effects of vocal intensity variation,'' J. Acoust. Soc. Am. **106**, 1532–1542.
Humes, L. E., Dirks, D. D., Bell, T. S., Ahlstrom, C., and Kincaid, G. E. (**1986**). ''Application of the articulation index and the speech transmission index to the recognition of speech by normal-hearing and hearing-impaired listeners,'' J. Speech Hear. Res. **29**, 447–462.
Johnson, K. (**1993**). ''The hyperspace effect: Phonetic targets are hyperarticulated,'' Language **69**, 505–528.
Krause, J. C., and Braida, L. D. (**2002**). ''Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility,'' J. Acoust. Soc. Am. **112**, 2165–2172.
Ladefoged, P. (**1975**). *A Course in Phonetics* (Harcourt Brace Jovanovich, New York).
Licklider, J. C. R., Hawley, M. E., and Walking, R. A. (**1955**). ''Influences of variations in speech intensity and other factors upon the speech spectrum,'' J. Acoust. Soc. Am. **27**, 207.
Miller, G. A., and Nicely, P. E. (**1955**). ''An analysis of perceptual confusions among some english consonants,'' J. Acoust. Soc. Am. **27**, 338–352.
Montgomery, A. A., and Edge, R. A. (**1988**). ''Evaluation of two speech enhancement techniques to improve intelligibility for hearing-impaired adults,'' J. Speech Hear. Res. **31**, 386–393.
Ochs, M. T., Humes, L. E., Ohde, R. N., and Grantham, D. W. (**1989**). ''Frequency discrimination ability and stop-consonant identification in normally hearing and hearing-impaired subjects,'' J. Speech Hear. Res. **32**, 133–142.
Payton, K. L., and Braida, L. D. (**1999**). ''A method to determine the speech transmission index from speech waveforms,'' J. Acoust. Soc. Am. **106**, 3637–3648.
Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). ''Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **95**, 1581–1592.
Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24**, 175–184.
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. **28**, 96–103.
Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). ''Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech,'' J. Speech Hear. Res. **29**, 434–446.
Pickett, J. M. (**1956**). ''Effects of vocal force on the intelligibility of speech sounds,'' J. Acoust. Soc. Am. **28**, 902–905.
Shannon, R. V., Zeng, F., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.
Summerfield, Q. (**1981**). ''Articulatory rate and perceptual constancy in phonetic perception,'' J. Exp. Psychol. Hum. Percept. Perform. **7**, 1074–1095.
Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., Schreiner, C., Jenkins, W. M., and Merzenich, M. M. (**1996**). ''Language comprehension in language-learning impaired children improved with acoustically modified speech,'' Science **271**, 81–84.
Tallal, P., and Piercy, M. (**1975**). ''Developmental aphasia: The perception of brief vowels and extended stop consonants,'' Neuropsychologia **13**, 69–74.
Turner, C. W., Smith, S. J., Aldridge, P. L., and Stewart, S. L. (**1997**). ''Formant transition duration and speech recognition in normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **101**, 2822–2825.
Uchanski, R. M., Choi, S., Braida, L. D., Reed, C. M., and Durlach, N. I. (**1996**). ''Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate,'' J. Speech Hear. Res. **39**, 494–509.

Performance of time- and frequency-domain binaural beamformers based on recorded signals from real rooms

Michael E. Lockwood,[a] Douglas L. Jones, Robert C. Bilger, Charissa R. Lansing, William D. O'Brien, Jr., Bruce C. Wheeler, and Albert S. Feng
Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, 405 North Mathews Ave., Urbana, Illinois 61801

(Received 2 November 2002; accepted for publication 22 August 2003)

Extraction of a target sound source amidst multiple interfering sound sources is difficult when there are fewer sensors than sources, as is the case for human listeners in the classic cocktail-party situation. This study compares the signal extraction performance of five algorithms using recordings of speech sources made with three different two-microphone arrays in three rooms of varying reverberation time. Test signals, consisting of two to five speech sources, were constructed for each room and array. The signals were processed with each algorithm, and the signal extraction performance was quantified by calculating the signal-to-noise ratio of the output. A frequency-domain minimum-variance distortionless-response beamformer outperformed the time-domain based Frost beamformer and generalized sidelobe canceler for all tests with two or more interfering sound sources, and performed comparably or better than the time-domain algorithms for tests with one interfering sound source. The frequency-domain minimum-variance algorithm offered performance comparable to that of the Peissig–Kollmeier binaural frequency-domain algorithm, but with much less distortion of the target signal. Comparisons were also made to a simple beamformer. In addition, computer simulations illustrate that, when processing speech signals, the chosen implementation of the frequency-domain minimum-variance technique adapts more quickly and accurately than time-domain techniques. [DOI: 10.1121/1.1624064]

PACS numbers: 43.72.Ew, 43.66.Pn, 43.66.Ts [DOS] Pages: 379–391

I. INTRODUCTION

A major problem for hearing aids, speech recognition, hands-free telephony, teleconferencing, and other acoustic processing applications is extracting, with good fidelity, a target sound in the presence of multiple competing sounds. This is particularly true of speech sounds, which are highly nonstationary in spectrum and intensity, and which may change position with respect to the listener over time. Thus, the cancellation of multiple, nonstationary, interfering speech sources requires fast, accurate tracking of the sources and robustness to reverberation and correlation between sources.

Many interference suppression techniques have been explored to address this problem. The most common approach is the use of an adaptive beamformer to process the sampled, time-domain outputs of a multimicrophone array, such that reception of the target sound from a particular direction is enhanced [see the reviews by Van Veen and Buckley, 1988; Brandstein and Ward, 2001]. These techniques use signals from the array to estimate the gradient of an error function and then iteratively move the filter coefficients closer to an optimal solution in small steps. Two algorithms that have been used extensively to address this problem are the iterative-adaptive techniques of Frost [1972] and Griffiths and Jim [1982]. Other variations of adaptive algorithms include those of Berghe and Wouters [1998] and Welker and Greenberg [1997]. Although these adaptive algorithms generally work well for suppressing statistically stationary interference sources that are uncorrelated with the target source, our experience has shown that they tend to adapt slowly or inaccurately in the presence of multiple, nonstationary interference sources such as speech, especially when there are more sources than sensors. As a result, the performance of the algorithms is compromised.

Greenberg and Zurek [1992] suggested that a solution to the problem of having more speech sources than sensors was to add more microphones. This solution is effective if all microphones are located far enough away from each other that they provide added useful inputs. However, accomplishing this in a hearing-aid system is difficult because locating microphones away from the ears is undesirable. Some current behind-the-ear (BTE) hearing aids contain two microphones per instrument; however, the small separation of the microphones limits the effectiveness of such systems. Likewise, it is impractical to use more than one microphone per instrument in systems that are located in the ear canal. Thus, a preferable system would use two sensors (one per instrument, located at each ear or in the ear canals), providing greater spatial separation of the microphones.

Several previous studies of adaptive beamformers [Greenberg and Zurek, 1992; Kompis and Dillier, 1994; Hoffman *et al.* 1994; Kates and Weiss, 1996] have avoided the problem of slow algorithm adaptation by allowing the adaptive filters sufficient time (at least 2 s for these studies) to converge before processing test signals. To provide more challenging test conditions, Greenberg *et al.* [2003] used a

[a]Electronic mail: melockwo@uiuc.edu

''roving'' interfering sound source that changed location at random times. However, this roving source was not used in all test conditions, and the algorithms in that study were permitted to converge for 1 s before the onset of the target signal. We argue that a beamforming algorithm used in a real hearing-aid instrument cannot be preadapted in an acoustically crowded environment due to the changing head position of the listener and the movement of the interfering talkers, and it may be unable to adapt quickly enough to perform effective interference suppression. Thus, the adaptation rate of a hearing-aid signal-processing algorithm is an important consideration.

To improve the adaptation rate of time-domain algorithms, an alternate approach to an iterative-adaptive technique is direct solution of the optimal beamformer [Capon, 1969]. While this in theory provides rapid convergence and improved interference cancellation, this technique generally requires the inversion of large, time-domain correlation matrices, a process that is inherently unstable and computationally impractical [Golub and Van Loan, 1996].

To suppress multiple, nonstationary interfering sources using only two sensors, frequency-domain beamforming algorithms appear to have distinct advantages compared to time-domain algorithms. For example, the algorithm of Liu *et al.* [1997, 2000, 2001] first determines source locations and strengths, and then performs constrained beamforming in each frequency band to remove interference. This method was shown to adapt quickly and could effectively suppress multiple speech interferers using only two microphones. It demonstrates a distinct improvement in performance over time-domain methods, but it requires intensive computation. The LENS algorithm [Deslodge, 1998] extracts a target signal by placing $n-1$ spatial nulls using n sensors. Greenberg *et al.* [2003] evaluated a four-sensor LENS implementation with up to three interfering sources.

Other frequency-domain algorithms do not perform beamforming, but rather attenuate individual frequency bands that contain interference. The algorithm of Peissig, Kollmeier, and colleagues (Kollmeier, 1997; Kollmeier *et al.*, 1993; Peissig and Kollmeier, 1997; Wittkop *et al.*, 1997), hereafter referred to as the P–K algorithm, uses coherence and phase and amplitude differences between channels to determine a gain for each frequency band. Slyh and Moses [1993] describe another example of this type of algorithm. These techniques have been shown to be effective in attenuating off-axis sources using only two sensors, but they also introduce signal distortion by attenuating part of the target signal.

Frequency-domain minimum-variance distortionless-response (MVDR) beamformers [Cox *et al.*, 1986; 1987] are more computationally efficient than both the technique of time-domain correlation matrix inversion and the algorithm of Liu *et al.* [2000, 2001]. MVDR beamformers pass signals from a target direction with no distortion, assuming the sensors are matched. Kates and Weiss [1996] used adaptive frequency-domain algorithms (preadapted for 2 s) to extract speech in a diffuse noise field using signals from a five-sensor end-fire array. Their study included an MVDR algorithm with a limited adaptation rate.

We hypothesized that a frequency-domain MVDR (FMV) algorithm, specifically implemented for fast adaptation, might be effective for suppressing multiple interfering speech sources using only two sensors. We have implemented such an algorithm and evaluated its performance with computer simulations [Lockwood, 1999; Lockwood *et al.*, 1999]. Initial tests showed that, compared to time-domain adaptive algorithms, the computational cost of the FMV algorithm is similar, but it converges much more quickly. For simulated signals with up to four interfering sources, the FMV algorithm outperformed the algorithms of Frost [1972] and Griffiths and Jim [1982] in terms of SNR gain [Yang *et al.*, 2000].

The focus of the current study is threefold. The first goal is to further evaluate the performance of the FMV algorithm with a two-sensor array in real environments. Although the performance of the FMV algorithm under simulated conditions is promising [Larsen *et al.*, 2001], it has not been evaluated under actual room conditions with real recorded signals. The second goal is to evaluate and compare time- and frequency-domain algorithms in acoustic scenes in which there are more speech sources than sensors. This represents a challenging condition for beamforming algorithms, and a condition that most studies have not explored. Only the study of Kates and Weiss [1996] evaluated algorithms in an environment with more sources than sensors, but the sources were all multitalker babble rather than speech. Finally, because speech sources and acoustic scenes change rapidly in real-world listening environments, the third goal is to develop a better understanding of how adaptation speed affects the performance of these algorithms.

Recordings were made in three different rooms with varying reverberation times (RTs: 0.10, 0.37, 0.65 s) using three different microphone arrays: (1) two microphones coupled to the ear canals of a KEMAR mannequin; (2) two omnidirectional microphones in free field separated by 15 cm; and (3) two cardioid microphones in free field separated by 15 cm. The performances of a two-channel FMV algorithm [Lockwood, 1999; Lockwood *et al.*, 1999], the Frost adaptive beamformer [Frost, 1972], a version of the generalized sidelobe canceler (GSC) [Greenberg, 1998; Griffiths and Jim, 1982], and an implementation of the Peissig–Kollmeier (P–K) [Kollmeier *et al.*, 1993; Wittkop *et al.*, 1997] binaural algorithm were assessed. The signals were also processed with a fixed beamformer. The algorithms were not allowed to preadapt, and their performance was compared in terms of the signal-to-noise ratio (SNR) gain and target signal distortion produced by each. The adaptation characteristics of the FMV, Frost, and GSC algorithms were further evaluated in computer simulation. Finally, the computational costs of the algorithms were examined.

II. ALGORITHM IMPLEMENTATIONS

A. Frequency-domain MVDR (FMV) algorithm

Time-domain input signals, with a sampling rate of 22.05 kHz, are transformed periodically (every $L=16$ samples) into the frequency domain via a length-N FFT, using a Hamming window. For a two-microphone system, the

TABLE I. Algorithm parameters for best performance.

Microphone	Processing algorithm: FMV	Frost	GSC	P–K
Omnidirectional (Sennheiser MKEII)	$N=1024$ $F=32$	$N_F=401$ $m_F=1.0$	$K_{GSC}=401$ $\alpha=0.15$	$N=1024$ $c_1=5$
	$M=1.03$	$c_F=0.01$		$c_2=1$ $c_3=1$
Cardioids (Sennheiser ME-104)	$N=1024$ $F=32$	$N_F=401$ $m_F=1.0$	$K_{GSC}=401$ $\alpha=0.15$	$N=1024$ $c_1=5$
	$M=1.03$	$c_F=0.01$		$c_2=1$ $c_3=1$
KEMAR (Etymotic ER-1)	$N=1024$ $F=32$	$N_F=401$ $m_F=1.0$	$K_{GSC}=401$ $\alpha=0.07$	$N=1024$ $c_1=1.0$
	$M=1.10$	$c_F=0.01$		$c_2=1.5$ $c_3=1.0$

frequency-domain signals from the sensors are represented by the components of the vector $\mathbf{X}_k=[X_{1k}\ X_{2k}]$, where k indexes the frequency bins. The F most recent FFTs are stored in a buffer, and a correlation matrix $\mathbf{R}_k$ is calculated for each frequency bin k by using:

$$\mathbf{R}_k=\begin{bmatrix} \frac{M}{F}\sum_{i=1}^{F} X_{1k,i}^{*}X_{1k,i} & \frac{1}{F}\sum_{i=1}^{F} X_{1k,i}^{*}X_{2k,i} \\ \frac{1}{F}\sum_{i=1}^{F} X_{2k,i}^{*}X_{1k,i} & \frac{M}{F}\sum_{i=1}^{F} X_{2k,i}^{*}X_{2k,i} \end{bmatrix}, \tag{1}$$

where * represents complex conjugation, and M is a multiplicative ''regularization'' constant slightly greater than 1.00 that helps avoid matrix singularity and improves robustness to sensor mismatch. Cox *et al.* [1987] described the use of additive regularization to control the trade-off between robustness and white-noise gain. Values of N, M, and F are found in Table I. The correlation matrices $\mathbf{R}_k$ are updated every $L=16$ samples, allowing them to quickly track changes of the input signals in all frequency bands. The correlation matrices and FFT buffers were set to zero before processing each signal.

For each frequency band k, the monaural output of the beamformer is

$$Y_k=\mathbf{w}_k^H\mathbf{X}_k, \tag{2}$$

where $\mathbf{w}_k$ is a vector of frequency-domain weights and ^{H} represents the Hermitian transpose of a vector. The optimization goal and constraint are expressed for each frequency band as

$$\min_{\mathbf{w}_k} E\{|Y_k|^2\}, \tag{3a}$$

$$\text{subject to}\quad \mathbf{e}^H\mathbf{w}_k=1, \tag{3b}$$

where min represents the minimization of a function with respect to selected variables (the weights, $\mathbf{w}_k$, in this case), $E\{\}$ represents the expected-value operation, and $\mathbf{e}$ is a vector indicating the desired arrival direction. This general approach is originally attributed to Capon [1969]. For an on-axis target source, both detectors receive the signal at the same time and with the same amplitude, assuming identical detectors. Thus, $\mathbf{e}^H=[1\ 1]$. If the desired receive direction were off-axis, then $\mathbf{e}$ would be complex valued. For the minimization goal and constraint given in Eqs. (3a) and (3b), an optimal solution is known [Capon, 1969; McDonough, 1979; Cox *et al.*, 1987]. For each frequency bin k, the optimal weight vector $\mathbf{w}_{\text{opt},k}$ is given by

$$\mathbf{w}_{\text{opt},k}=\frac{\mathbf{R}_k^{-1}\mathbf{e}}{\mathbf{e}^H\mathbf{R}_k^{-1}\mathbf{e}+\sigma}, \tag{4}$$

where $\mathbf{R}_k$ is defined in Eq. (1), $\mathbf{R}_k^{-1}$ represents the matrix inverse of $\mathbf{R}_k$, and σ is a very small positive quantity that prevents division by zero. Inherent to this solution is the assumption that it is valid only if the inputs are stationary random processes. This is assumed to be true for small time intervals of speech signals in each frequency band.

To respond quickly to changes in $\mathbf{R}_k$, new optimal weights $\mathbf{w}_k$ are calculated for half of the frequency bands every L samples, so all weights are updated every $2L$ samples. This is possible because for a two-sensor system, the matrix inversion for each frequency band is computationally inexpensive. It will be demonstrated in Sec. V that this technique yields faster and more accurate tracking of nonstationary sources than time-domain techniques.

To obtain the time-domain output, the newest optimal weights for each frequency band are applied to buffered FFT data to obtain the output [Eq. (2)]. The resulting N frequency-domain values are then transformed to the time domain using a length-N inverse FFT. This occurs every L samples, and the central L samples of time-domain output are used. As the outer samples of the FFT window are attenuated by the Hamming window, this minimizes the effects of circular convolution which arise due to the FFT-based filtering, while requiring less computation and delay than an overlap-save or overlap-add method [Joho and Moschytz, 2000].

The main consideration in the FMV implementation was frequency resolution, which is determined by the FFT length relative to the sampling rate. For our experiments, a 1024-point FFT was chosen because it provided the best performance. Increasing the FFT length decreases the bandwidth of each frequency bin and should improve FMV performance (for stationary signals), as this provides the more detailed estimates of signal spectra. However, in practice, when the FFT length was too long the performance decreased, likely because the signal was not sufficiently stationary for the interval of the FFT. Also, longer FFTs required more data points, and objectionably increased the system delay.

B. Frost, GSC, P–K, and fixed beamformer implementations

Optimized parameter values (see the next section for details) for three of the algorithms described in this section are shown in Table I. All time-domain adaptive weights were initialized to that of a conventional beamformer, with a value of 0.5 for each channel at an appropriate delay, and zero

otherwise. All time-domain weights were updated each sample. The correlation matrices for the P–K algorithm were also set to zero before processing and were updated every 16 samples (as with the FMV algorithm).

The Frost algorithm was implemented with the update equation

$$\mathbf{W}_{\text{new}} = \mathbf{P} \cdot \left(\mathbf{W}_{\text{old}} - \frac{2}{3} \cdot \frac{\mathbf{W}_{\text{old}}^T \cdot \mathbf{x}_f \cdot \mathbf{x}_f}{m_F \cdot \mathbf{x}_f^T \cdot \mathbf{x}_f + c_F} \right) + \mathbf{F}, \quad (5)$$

where $\mathbf{W}_{\text{old}}$ is the previous set of time-domain filter coefficients, and m_F and c_F are adjustable parameters to control the step size and to prevent divide-by-zero, respectively. Additionally, $\mathbf{x}_f$ is a column vector composed of the time-domain input signals from both channels, $\mathbf{P}$ is a precomputed projection matrix, and $\mathbf{F}$ represents the response constraints, all as per Frost [1972]. The factor of 2/3 facilitates comparison of the step size with a bound derived by Frost. N_F is defined as the length of the adaptive filter.

The generalized sidelobe canceler (GSC) algorithm [Griffiths and Jim, 1982] was implemented as per Greenberg [1998]. This implementation improves performance and reduces target distortion in nonstationary environments when the target signal is strong. The weight update equation was

$$\mathbf{W}_{\text{new}} = \mathbf{W}_{\text{old}} + \frac{\alpha_{\text{sum}}}{K_{\text{GSC}}[\sigma_e^2(n) + \sigma_x^2(n)]} \cdot e(n) \cdot \mathbf{x}_G(n), \quad (6)$$

where α_{sum} is a step size parameter, n is an index of the current sample, $\mathbf{W}_{\text{old}}$ is the previous set of time-domain filter coefficients, e is the processed output, $\mathbf{x}_G$ is a vector of samples of the signal passed by the blocking matrix (mostly interference), K_{GSC} is the filter length, and σ_e^2 and σ_x^2 represent the average powers (updated every sample) of e and $\mathbf{x}_G$, respectively.

The P–K algorithm [Kollmeier *et al.*, 1993; Wittkop *et al.*, 1997] was implemented with three variable parameters to control the attenuation of the signal as a function of the phase and amplitude differences between channels and the coherence between channels. For the kth frequency band, the (real-valued) filter weight G_k was determined using

$$\begin{aligned} G_k &= g_{1k} \cdot g_{2k} \cdot g_{3k}, \\ g_{1k} &= \max\left(0,\ 1 - \frac{c_1 |\angle R_{12}|}{\pi}\right), \\ g_{2k} &= c_2 \cdot \min\left(\frac{R_{11}}{R_{22}}, \frac{R_{22}}{R_{11}}\right), \quad g_{3k} = \left(\frac{\text{Re}[R_{12}^2]}{R_{11} \cdot R_{22}}\right)^{C_3}, \end{aligned} \quad (7)$$

where c_1, c_2, and c_3 are adjustable parameters that control the sensitivity of the algorithm to phase differences between channels, amplitude differences between channels, and coherence between channels, respectively; $\angle$ represents the phase angle in radians, Re[ ] represents the real part of a complex value, and R_{ij} is an element from a frequency-domain correlation matrix [see Eq. (1)], with $M = 1.00$.

A simple fixed beamformer (referred to as a conventional beamformer) was implemented by averaging the signals from the two microphones after a matching filter was applied.

C. Optimization

All algorithms and metrics were implemented using MATLAB 6.5 (The MathWorks, Inc., Natick, MA). Floating-point calculations were performed with 64-bit precision. The algorithm parameters (Table I) were adjusted for best performance in terms of the SNR metric [Sec. III F, Eq. (9)]. The test signals were processed with many different sets of parameters. For the frequency-domain algorithms, the effect of changing the FFT length was generally independent of the effects of changing other parameters, so this was set first. Additionally, the additive constant c_F in the Frost algorithm was found to have little effect on performance so long as it was above a minimum value. This narrowed the parameter space to two variables for the FMV, Frost, and GSC algorithms, and to three variables for the P–K algorithm.

Many sets of values were chosen for the remaining free parameters, and the signals were processed and results obtained for all. It was found that performance varied with the number of interfering sources. Because this study emphasizes the effects of multiple interfering sources (more sources than sensors), the parameter set for each algorithm that produced the best overall performance for tests with multiple interferers was chosen as optimal. This was an *ad hoc* decision made by plotting the performance on a graph and choosing the line that was highest for the multiple-interferer tests. For the time-domain algorithms, special care was taken not to choose parameter sets that caused instability, as it was possible to have good multiple-interferer performance while the algorithm became unstable for the one-interferer test signals.

III. EXPERIMENTAL METHODS

A. Test materials

A series of high-context sentences by eight talkers (four females and four males) from the revised R-SPIN test [Bilger *et al.*, 1984] were recorded on digital audio tape (DAT) at a sampling rate of 48 kHz, quantized to 16 bits. Recordings were made in a sound-treated studio (model: Studio 7×5×5 ft., Acoustic Systems, Austin, TX). The recorded sentences were downsampled to 44.1 kHz. Three different sentences were chosen from each talker. This provided a total of 24 different sentences, 12 by male talkers and 12 by female talkers. A section of multitalker babble from the R-SPIN test was also used as an interfering signal in several test configurations. Its sampling rate was also 44.1 kHz.

B. Recording techniques and setup

Each of the 24 sentences, the multitalker babble signal, and 10 s of white noise were played back from eight loudspeakers housed in a semicircular enclosure (SPATS, Sensimetrics Corp., Somerville, MA). Each loudspeaker was equidistant (75 cm) from a central point and comprised a single 7.6-cm-diameter driver with frequency response restricted to 200-Hz to ~13 kHz. The two-microphone arrays were located at the central point of the array, 1.15 m above the floor,

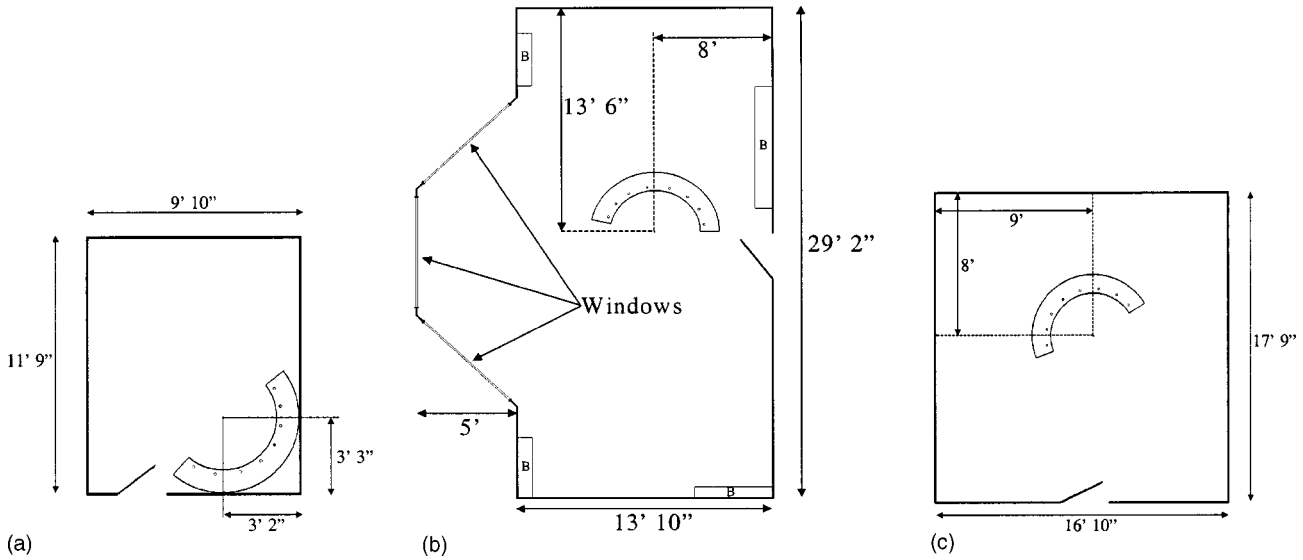


FIG. 1. (A) Diagram of room 1 (treated with acoustic foam) with loudspeaker array. (B) Diagram of room 2 (conference room with windows) with loudspeaker array. (C) Diagram of room 3 (conference room, bare walls) with loudspeaker array. Filled circle represents target loudspeaker, open circle represents interferer; distances to central point of the array are shown.

and oriented such that the loudspeakers were at azimuths of 60°, 40°, 20°, 0°, −20°, −40°, −60°, −80° with respect to the broadside array.

Recordings were made with three sets of microphones: (1) Sennheiser MKEII omnidirectional microphones spaced 15 cm apart in free field; (2) Sennheiser ME104 cardioid microphones spaced 15 cm apart in free field; and (3) Etymotic ER-1 microphones mounted in the ears of a KEMAR mannequin (Knowles Electronics, Itasca, IL). The omnidirectional and cardioid microphones were connected directly to a microphone preamplifier (Millennia Media HV-3B, Placerville, CA) that was connected to the inputs of the Aark 24 system. The KEMAR microphones were connected to their own dedicated preamps, and then to the Millennia Media preamplifier. Recording and playback were done with a sampling rate of 44.1 kHz. The recorded data were downsampled to 22.05 kHz prior to being saved to hard disk.

C. Room descriptions

Recordings were made in three rooms with different reverberation characteristics. In all rooms, the ceiling had acoustic tile suspended at a height of 9 ft. Above it was a concrete ceiling. The floors of all rooms were covered with short carpet. The dimensions of the rooms and the positioning of the loudspeaker and microphone arrays within them are shown in Figs. 1(a), (b), and (c). Room 1 was an acoustically controlled space with 6.4-cm-thick foam (SONEX Valueline, Illbruck Corp., Minneapolis, MN) attached to all wall surfaces; Room 2 was a rectangular conference room with bay windows and bookshelves on two walls; Room 3 was a rectangular conference room with painted gypsumboard walls.

The reverberation times of the rooms were measured with a sound-level meter (Bruel & Kjaer, model 2260). A more suitable source of an impulse was not available, so a hand clap was used as a stimulus after it was found that it provided consistent results over several measurements. Five measurements were taken in each room at the central point of the loudspeaker array and the results were averaged. The meter calculated the T_{60} times for 1/3-octave bands between 200-Hz and 10 kHz. The values for each band were averaged over the five measurements, and then averaged across frequency bands to obtain the reverberation time. The estimated average T_{60} values for rooms 1, 2, and 3 were 0.10, 0.37, and 0.65 s, respectively.

D. Compensation, energy equalization, test creation

To match the microphones, the recorded white noise from the loudspeaker at 0° was filtered with a 43-tap FIR filter adapted by an LMS algorithm to match the responses of the microphones for the target direction. For a sampling rate of 22.05 kHz, this filter has a 2-ms delay, corresponding to a sound propagation distance of approximately 0.7 m. Given the dimension of the rooms and location of the microphones, direct reflections should require more than 2 ms to reach the microphones; the filters should thus compensate mostly for the differences in frequency response of the microphones, rather than responding to reflections from the room.

All acoustic recordings (except for the recordings of white noise) were of the same duration; however, the duration of the spoken sentences varied. Therefore, the average power of each recorded sentence was calculated (taking into account the varying durations) and all sentences were scaled to have the same average power. This is referred to as the normalized energy of a single interfering source. The amplitude of each interfering source could then be adjusted to provide a specified relative amount of interference with respect to the target sentence.

To construct a test, the recordings of each source from its appropriate location (loudspeaker) were scaled and added, keeping the left and right channels separated, thereby producing a binaural test signal. This assumed that the acoustic addition of the sources and the various components in the recording and playback system all act in a linear manner.

TABLE II. Summary of test configurations.

Configuration number	Azimuth angle of source				
	+60°	+20°	0°	−40°	−80°
1	Interferer[a]		Target		
2		Interferer	Target		
3	MTB[a]		Target		
4		MTB	Target		
5	Interferer		Target		Interferer
6	Interferer		Target	Interferer	
7		Interferer	Target	Interferer	
8	Interferer	Interferer	Target		Interferer
9	Interferer	Interferer	Target	Interferer	
10	Interferer	Interferer	Target	Interferer	Interferer
11	Interferer or MTB	Interferer or MTB	Target	Interferer or MTB	Interferer or MTB

[a]"Interferer" and "MTB" indicate interference from the angle listed by a single talker and multitalker babble source, respectively.

E. Test terminology and guidelines

A test configuration was defined as one particular spatial arrangement of target and interfering sources. Eleven test configurations were used; Table II describes the positioning of the target and interferer(s). Each configuration contained one on-axis target source (always a single talker) presented from the loudspeaker at 0°, and one to four off-axis interfering sources presented from the loudspeakers at possible azimuth angles of +60°, +20°, −40°, or −80°, each a single talker or multitalker babble source. Multitalker babble was used in test configurations 3 and 4 as the only interferer. For test configuration 11, a multitalker babble was placed at one of the four interference locations and single-talker interferers were placed at the other three.

All sources were located in the front half-plane, and thus the effects of the directionality of the cardioid microphones was minimal. Additionally, a front–back ambiguity did not exist as there was no source located greater than 90° from the target signal.

For each spatial test configuration, 12 tests were constructed. A test was defined as one permutation of a target sentence and interfering sentence(s). Twelve different sentences, three sentences each from two male and two female talkers, were used exclusively as target sentences. (Thus, for each test configuration, six target talkers were males and six were females.) The remaining 12 sentences were used exclusively as interferers. Additionally, in any test, each source was a different talker, and all interferers had the same energy.

Each test was constructed and processed at three different SNRs. For configurations 1–4 (one interferer), three normalized energies were used: −3, 0, and +3 dB (corresponding to each interference source having 3 dB less, the same, or 3 dB more average power, respectively, than the target source). For test configurations 5–11, three lower normalized energies (−6, −3, 0 dB) were used. Therefore, the entire battery of tests consisted of 11 test configurations, 12 tests per configuration, and 3 SNR levels per test, for a total of 396 test signals. Each test was approximately 2.5 s long, for a total of 16.5 minutes of test signals. Results for each test configuration are presented in Sec. IV, and were obtained by averaging the performance metrics over the 12 tests and three SNR levels for each configuration.

F. Performance metrics

Processing with the various algorithms was done off-line and the signals from the individual sources before and after processing were known, greatly simplifying the calculation of performance metrics. Because the target signal could be distorted by significant amounts during processing, and because distortion is detrimental to speech perception, a signal-to-noise ratio (SNR) metric that incorporated both interference and signal bias (distortion) error was chosen. The output SNR (after processing) is defined as

$$\mathrm{SNR}_{\mathrm{OUT}} = 10 \cdot \log_{10}\left(\frac{\Sigma_{v=1}^{V} t_u(v)^2}{\Sigma_{v=1}^{V} (y_p(v) - t_u(v))^2} \right), \tag{8}$$

where $y_p(v)$ and $t_u(v)$ are the vth samples of the processed output and unprocessed (ideal) target signals, respectively, and V is the length of the signal in samples. In this work, the unprocessed signal is binaural and the output is monaural. Thus, the expression is modified to be

$$\mathrm{SNR}_{\mathrm{OUT}} = 10 \cdot \log_{10}\left(\frac{\Sigma_{i=1}^{V} [g_1 t_{u,L}(v) + g_2 t_{u,R}(v)]^2}{\Sigma_{i=1}^{V} (y_p(v) - (g_1 t_{u,L}(v) + g_2 t_{u,R}(v)))^2} \right), \tag{9}$$

where $t_{u,L}(v)$ and $t_{u,R}(v)$ are the unprocessed target signals from the left and right microphones, respectively, and g_1 and g_2 represent the gains applied by the algorithm (effectively the steering vector) to the target signal in each channel. In this study, g_1, $g_2 = 0.5$ because filters are used to match the responses of the microphones in the target direction (see Sec. III D). If the microphones are not matched for the target direction, the gains must be replaced with filters.

The input SNR for one channel is defined as

$$\mathrm{SNR}_{\mathrm{IN}} = 10 \cdot \log_{10}\left(\frac{\Sigma_{v=1}^{V} t_{u,L}(v)^2}{\Sigma_{v=1}^{V} i_{u,L}(v)^2} \right), \tag{10}$$

where $i_{u,L}(v)$ is the interference signal received by the left microphone. Because all signals were recorded individually, the target and interference signals received by the micro-

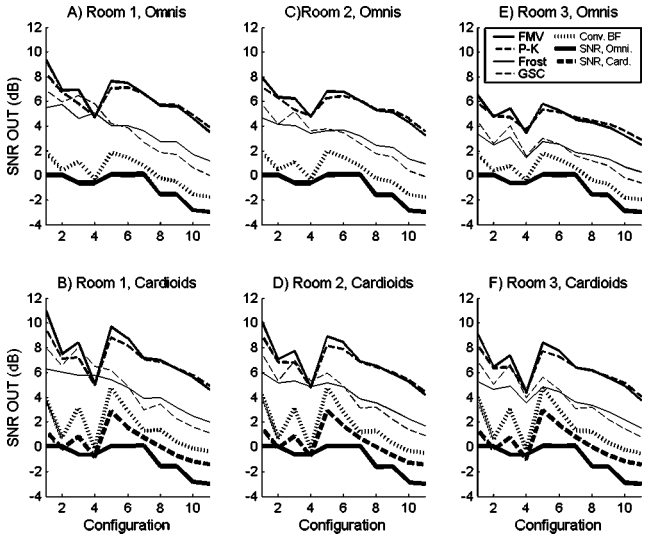


FIG. 2. (A), (C), (E) Processed and unprocessed SNR values for the left omnidirectional microphone data in rooms 1, 2, and 3, respectively. (B), (D), (F) Processed and unprocessed SNR values for the left cardioid microphone data in rooms 1, 2, and 3, respectively.

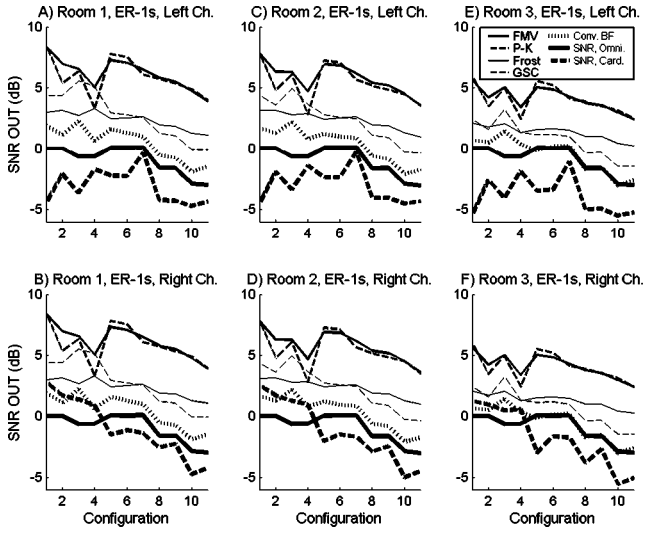


FIG. 3. (A), (C), (E) Processed and unprocessed SNR values for left KEMAR microphone data in rooms 1, 2, and 3, respectively. (B), (D), (F) Processed and unprocessed SNR values for right KEMAR microphone data in rooms 1, 2, and 3, respectively.

phones are known and the SNR_{IN} calculation yields the SNR after microphone reception.

A second metric used was a measure of the target distortion. Access to the processed target signal allowed this distortion to be quantified. The FMV, Frost, and GSC algorithms are constrained to pass the target with zero distortion, and are referred to as distortionless response beamformers. However, microphone mismatch and reverberation mean that only a portion of the target signal satisfies the constraint, so target distortion may occur. The P–K algorithm attenuates frequency bands that contain interference, and thus almost always cancels some of the target signal. We define a measure of target distortion as

$$T_D = \frac{\Sigma_{v=1}^{V}(t_p(v) - [g_1 t_{u,L}(v) + g_2 t_{u,R}(v)])^2}{\Sigma_{v=1}^{V}[g_1 t_{u,L}(v) + g_2 t_{u,R}(v)]^2}, \qquad (11)$$

where $t_p(v)$ is the processed (monaural) target signal. When T_D is greater than zero, the target signal has been distorted. Note that this metric does not distinguish between attenuation and phase distortion.

IV. RESULTS AND DISCUSSION

A. Algorithm comparisons

Figure 2 (results for omnidirectional and cardioid microphones, left channel) and Fig. 3 (results for KEMAR microphones, both channels) show the average SNR_{OUT} produced by the algorithms after processing. The SNR_{IN}, as received by the omnidirectional, cardioid, and KEMAR (ER-1) microphones, is labeled as (SNR, omni), (SNR, card), and (SNR, ER-1), respectively. To facilitate the comparisons in Sec. IV C, (SNR, omni) is also shown with the results for the cardioid and KEMAR microphones. In all three rooms and for all microphone types, the FMV and P–K algorithms consistently outperform the Frost and GSC algorithms in terms of SNR_{OUT} for all test configurations with more than one interfering source (configurations 5–11). This is the main advantage conferred by the frequency-domain algorithms.

For the one-interferer tests (configurations 1–4), the FMV and P–K algorithms perform better than the time-domain algorithms in terms of SNR_{OUT} for configurations 1 and 2. For configurations 3 and 4, the time-domain algorithms perform similarly or slightly better in room 1, and in room 2 with cardioid microphones. (Though the algorithm parameters were not optimized for these one-interferer tests, the improvement in performance that can be obtained by optimizing the parameters for these tests is not very large.)

The FMV and P–K algorithms perform similarly in

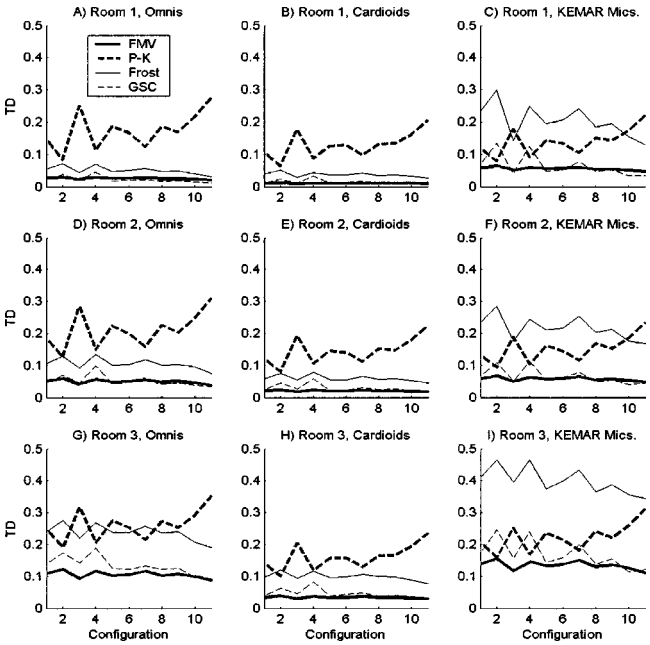


FIG. 4. (A), (D), (G) Target distortion for omnidirectional microphones in rooms 1, 2, and 3, respectively. (B), (E), (H) Target distortion from cardioid microphones in rooms 1, 2, and 3, respectively. (C), (F), (I) Target distortion for KEMAR microphones in rooms 1, 2, and 3, respectively.

terms of SNR_{OUT}, but the P–K produces more target distortion, as can be seen in Fig. 4. This is to be expected, as the FMV algorithm uses constrained spatial filtering to remove interference without distorting the target signal (with matched sensors and no reverberation), while the P–K algorithm simply attenuates frequency bands that appear to be dominated by interference. Therefore, target distortion is accepted in exchange for an improved SNR_{OUT}.

The differences between the GSC and the Frost algorithm are generally small. Griffiths and Jim [1982] showed that their GSC structure converges to the same solution as the Frost beamformer, although their paths to convergence may be different. In the tests conducted here, the GSC algorithm appears to have a slight performance advantage over the Frost algorithm for the single-interferer tests, while the Frost algorithm appears to perform better than the GSC as the number of interferers rises. This difference is more apparent with the KEMAR microphones (Fig. 3), for which the GSC has a more pronounced advantage in the one-interferer tests (up to 3 dB) and for which the Frost algorithm performs better in the multiple-interferer tests (by up to 2 dB). The cause of these differences is not known.

The FMV and GSC algorithms generally produce less distortion (Fig. 4) of the target signal than the Frost and P–K algorithms. The FMV and GSC distortion figures are often comparable, but the FMV has an advantage in the more reverberant environments. The high distortion figures of the P–K algorithm are expected, but those for the Frost algorithm are not. They suggest that the Frost algorithm is more sensitive to the amount of reverberation and the type of microphone being used than the GSC or FMV algorithms are.

The conventional beamformer, which averages the two inputs, performs more poorly than all other algorithms, as shown in Figs. 2 and 3.

B. Effects of reverberation

The performance, in terms of SNR_{OUT} (Figs. 2 and 3), of all algorithms generally decreases by 1 dB or less for the tests in room 2 ($RT=0.37$ s) as compared to room 1 ($RT=0.10$ s). This decrease is fairly consistent across the different types of microphones. Performance differences between room 2 and room 3 ($RT=0.65$ s) are slightly more pronounced, including a drop of 1–2 dB for the FMV and P–K algorithms across all configurations. Frost and GSC algorithm performance is reduced by a similar amount.

In general, the performance of all algorithms is decreased by similar amounts as the reverberation time of the room increases. Importantly, this implies that the performance advantage of the frequency-domain algorithms is maintained as the reverberation time increases. For the one-interferer configurations, the advantage of the frequency-domain algorithms actually increases with reverberation time; this implies that the advantages of these algorithms are not limited to the configurations in which there are more sources than sensors.

It is important to point out that distortion for distortionless-response beamformers (FMV, Frost, GSC) is due to sensor mismatch (including mis-steering) and reverberation. The matching filters were chosen carefully, but they were kept sufficiently short so that they did not compensate for reflections and reverberation. Therefore, the distortion figures generally reveal how sensitive the algorithms are to reverberation.

Increasing the amount of reverberation consistently increases target distortion. Figure 4 reveals that the target distortion for the FMV, Frost, and GSC algorithms approximately doubles from room 1 to room 2, and doubles again from room 2 to room 3. The FMV and GSC generally have the lowest distortion for all rooms. The distortion of the P–K algorithm rises by approximately 0.05 for both rooms 2 and 3. The P–K algorithm has by far the highest distortion in rooms 1 and 2 with omnidirectional and cardioid microphones; Frost has the second highest. The distortion of the Frost algorithm is comparable to that of the P–K algorithm for room 3. While the FMV, Frost, and GSC algorithms appear to be more sensitive to increases in reverberation time than the P–K algorithm, for the range of RTs used in this study, the distortion of the FMV and GSC algorithms is still notably less than that of the P–K algorithm.

C. Microphone effects

The directionality of the cardioid microphones accounts for up to a 3-dB improvement in the SNR_{IN} (SNR, card) over that of the omnidirectional microphones (SNR, omni), as can be seen in Fig. 2. This appears to account for the approximately 1–2-dB improvement in the SNR_{OUT} that is observed for all algorithms for the multiple-interferer tests in all three rooms. Overall, the FMV algorithm with cardioid microphones performs best, albeit by a small margin.

The KEMAR microphones (Fig. 3) cause a reduction in the SNR_{IN} for the left channel and an increase in the SNR_{IN} (SNR, ER-1) for the right channel for single-interferer tests. This is because the interfering source for these tests is from the left of the array, at $+60°$ or $+20°$, and thus the interference is stronger in the left ear of the KEMAR than the right. For multiple interferer tests, the SNR_{IN} is lower than with the other microphones. Thus, processing the KEMAR signals yields 1–2-dB lower SNR_{OUT} than with omnidirectional or cardioid microphones. Another notable effect of the KEMAR microphones is the dramatic increase in distortion [Figs. 4(c),(f),(i)] for the Frost algorithm, likely caused by sensitivity to reverberation or microphone mismatch. The distortion for the GSC algorithm remains low.

For the experiments in this study, all sound sources were placed in the front half-plane. This reduced the benefits obtained by using the cardioid or KEMAR microphones. If sources had been placed in the back half-plane, using these microphones would have resulted in higher values of SNR_{IN}. However, for this study, the effect of the directionality of the microphones that is most interesting is the ability to reduce the reverberation in the recorded signals. It was hoped that the directional microphones would make all algorithms more robust to reverberation effects. Compared to the processed signals from the omnidirectional microphones, processing the signals from the cardioid microphones produces a somewhat higher output SNR (for all three rooms) and generally lower target distortion, while processing signals from the

TABLE III. Frequencies of sinusoidal interferers and positions as a function of time.

Time interval	Sinusoid frequencies (Hz)									
	500	611	682	769	921	1016	1095	1187	1331	1448
	Azimuth angles									
0.00–0.75 s	−65°	−55°	−45°	−35°	−25°	20°	30°	40°	50°	60°
0.75–1.50 s	20°	30°	40°	50°	60°	−65°	−55°	−45°	−35°	−25°
1.50–2.25 s	60°	50°	40°	30°	20°	−25°	−35°	−45°	−55°	−65°

KEMAR microphones produces slightly lower SNRs and higher distortion. Overall, the performances of the algorithms were improved by the use of the cardioid directional microphones, and only slightly reduced by the KEMAR microphones.

D. Best parameter sets

The optimal parameter sets for each algorithm (Table I) were found to be the same for both omnidirectional and cardioid microphones, but different for the KEMAR microphones. As compared to the best parameter sets for the free-field microphones, the FMV algorithm performs best with a larger regularization value M, the GSC algorithm requires a smaller step size α to remain stable in multiple interferer tests, and the P–K algorithm performs best with values of c_1 and c_2 that weight the amplitude difference most heavily instead of the phase difference.

V. TIME-VERSUS FREQUENCY-DOMAIN MVDR BEAMFORMERS

A. Overview

The results in Sec. IV show the performance advantage of the FMV over conventional time-domain algorithms. The Frost and GSC algorithms differ from the FMV in that they are iterative-adaptive time-domain techniques, while the FMV calculates optimal solutions directly for many frequency bands. However, all are classified as MVDR beamformers, because they have identical optimization goals and constraints (minimize output power, pass target source undistorted). These MVDR beamformers (time- and frequency-domain) will all asymptotically converge to identical solutions for inputs that are stationary random processes. Therefore, differences in SNR performance are due to different adaptation characteristics in the presence of nonstationary signals such as speech.

Using specific simulated signals, we now examine the reasons for the performance advantage of the FMV over the other algorithms. The test signals are not true binaural recordings as in the previous experiments, rather they are simulated. The signals for each source in each channel differ only by time delay. Therefore, no sensor mismatch is present, the FMV, Frost, and GSC algorithms produce no target distortion, and performance differences result only from the varying adaptation characteristics of the algorithms. The absence of sensor mismatch also allows accurate beam patterns to be calculated to observe these characteristics.

The algorithms were all optimized to produce the best SNR gain for the simulated signals. The first example illustrates the behavior of MVDR beamformers with statistically stationary interference. The second example shows that the FMV is able to adapt more quickly and accurately to rapid changes in interfering sound sources, and thus it more effectively attenuates multiple-interfering sound sources.

B. Simulation example 1: Multiple spatially separated sinusoidal interferers

MVDR beamforming algorithms may, in theory, attenuate multiple, statistically stationary narrow-band interference signals at different frequencies. To demonstrate this, a two-channel test signal was synthesized containing a single on-axis speech source and ten off-axis sinusoidal signals, each with a different frequency and azimuth. The sinusoid frequencies ranged from 500 to 1450 Hz, with approximately 100-Hz spacing. Neglecting frequency smearing due to win-

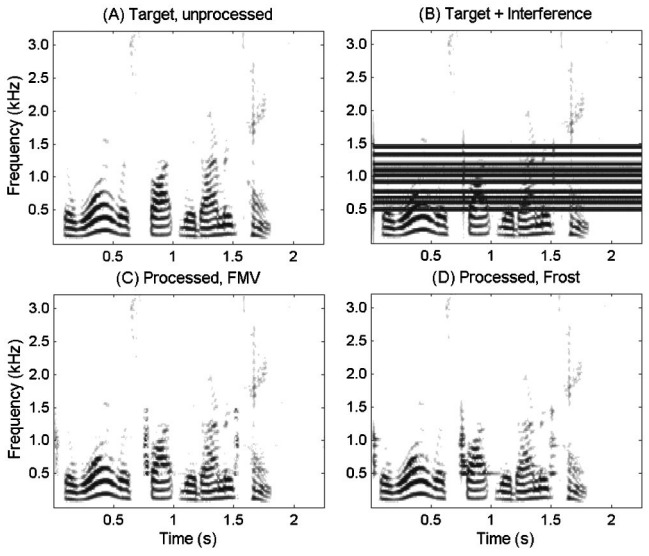


FIG. 5. (A) Spectrogram of the target speech signal in example 1 (''The war was fought with armored tanks.'') before processing. (B) Spectrogram of the combined target and sinusoidal interference signals in example 1 before processing. The target signal is at 0° azimuth, and ten sinusoidal interferers originate from angles between ±65° azimuth. (C), (D) Spectrograms of example 1 after processing with the FMV algorithm (C) and the Frost algorithm (D).

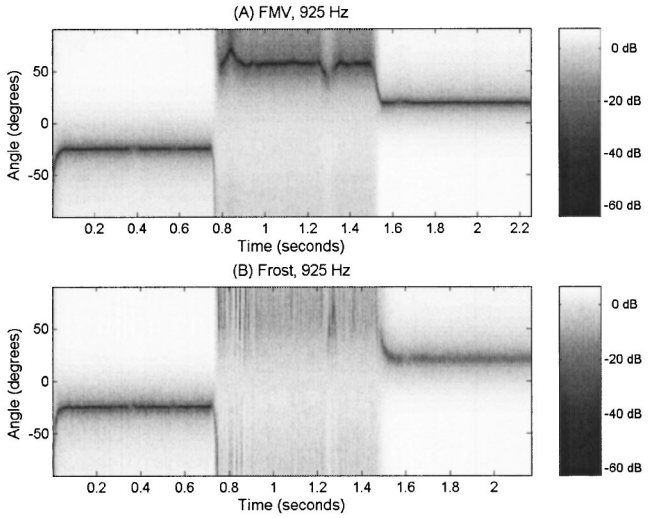


FIG. 6. (A) Magnitude (in decibels) of the beam pattern for the FMV algorithm for the 925-Hz FFT bin. Note the null at $-25°$ ($t=0.00$ to 0.75 s), $\approx 60°$ ($t=0.75$ to 1.50 s), and 20° ($t=1.50$ to 2.25 s). (B) Magnitude (in decibels) of the beam pattern for the Frost algorithm for the 925-Hz FFT bin. The null location varies considerably, but on average is near $-25°$ ($t=0.00$ to 0.75 s), 60° ($t=0.75$ to 1.50 s), and 20° ($t=1.50$ to 2.25 s).

dowing, there is no spatial or frequency overlap of the interferers. So that the adaptation rate of both algorithms can be observed, the sinusoids instantaneously change locations at 0.75 and 1.5 s. The sinusoid frequencies and azimuths as a function of time are listed in Table III.

Spectrograms of the target and combined signals before processing are shown in Figs. 5(a) and (b), respectively. The SNR before processing is -7.35 dB. The spectrograms of the processed output from the FMV and Frost algorithms are shown in Figs. 5(c) and (d), respectively, and the SNR gains were 18.56 and 16.14 dB, respectively.

While processing this signal, all of the frequency-domain weights calculated by the FMV algorithm for the FFT bin centered on 925 Hz were recorded. All of the time-domain filter coefficients from the Frost algorithm were transformed to frequency-domain coefficients and recorded. Using these coefficients, the beam pattern for the 925-Hz FFT bin (the gain, in decibels, for signals as a function of azimuth angle and time) was calculated for both algorithms, and is shown in Fig. 6(a). A two-element beamformer (with a constraint) can place at most one spatial null per frequency band, and the null direction changes with time as it adapts to the input signal. The null is easily visible, occurring at $-25°$, near 60°, and at 20° for time intervals [0.0, 0.75 s], [0.75, 1.5 s], and [1.5, 2.25 s], respectively. Table III shows that a sinusoidal interferer with a frequency of 921 Hz was at these locations at the above time intervals.

Figure 6(a) also shows that the FMV algorithm null deviates from the direction of the interferer between 0.75 and 1.5 s. Examining Fig. 5(b) reveals that the spectra of the target (speech) and interference (sinusoidal) sources overlap in the 925-Hz FFT bin between 0.75 and 1.5 s. The optimal filter weights are calculated assuming that there is no correlation between target and interfering sources; in practice, this is rarely true over short time intervals. Thus, the correlation matrix estimates contain some error. (To reduce this error, correlation matrices may be formed using data from longer time intervals with the trade-off of slower adaptation.)

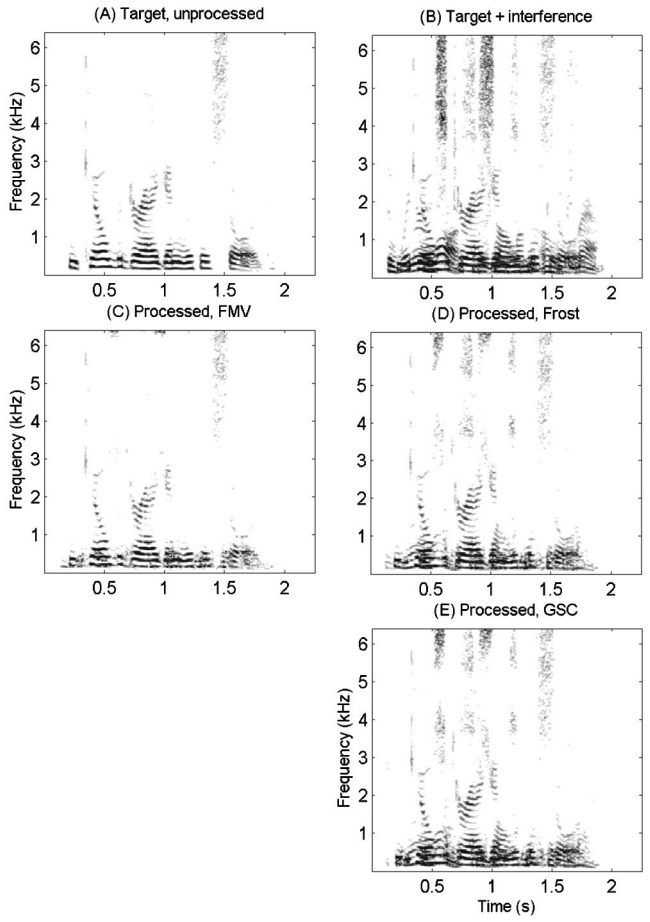


FIG. 7. (A) Spectrogram of target speech signal in example 2 (''He killed the dragon with his sword.'') before processing. (B) Spectrogram of combined target and two interfering speech signals in example 2, before processing. The target signal is at 0° azimuth and the interference sources are at $\pm 45°$. (C), (D), (E) Spectrogram of example 2 after processing with the FMV (C), Frost (D), and GSC (E) algorithms.

Figure 6(b) shows the beam pattern for the Frost algorithm for the 925-Hz FFT bin. The direction of the null placed by the Frost algorithm is erratic when the signals overlap (between 0.75 and 1.5 s). This occurs because a noisy time-domain estimate of the gradient is used to adapt the filter weights. This misadjustment error can be reduced by lowering the adaptation rate, but only at the expense of slower convergence, thus leading to poorer performance in the presence of multiple interferers.

In this example the FMV and Frost algorithms show nonideal effects when target and interference signals overlap. However, both algorithms effectively attenuate multiple, nonoverlapping statistically stationary interferers using signals from two sensors.

C. Simulation example 2: Multiple speech interferers

This example uses statistically nonstationary interference sources (speech) to compare the speed and accuracy of convergence of the FMV, Frost, and GSC algorithms. An on-axis speech source and two off-axis interfering speech sources (at $\pm 45°$) are simulated. Figures 7(a) and (b) show

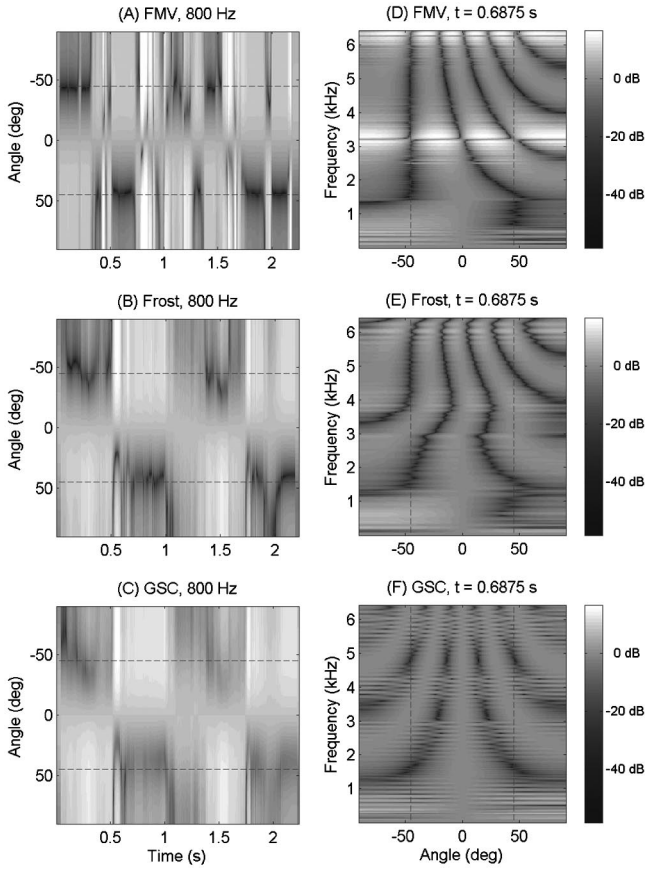


FIG. 8. (A), (B), (C) Magnitude (in decibels) of the beam patterns for the FMV (A), Frost (B), and GSC (C) algorithms for the 800-Hz FFT bin. Nulls near $\pm 45°$ correspond to cancellation of interfering speech sources at $\pm 45°$. (D), (E), (F) Magnitude (in decibels) of the beam patterns at $t = 0.6875$ s for the FMV (D), Frost (E), and GSC (F) algorithms.

the spectrograms of the target and target plus interference signals, respectively. The SNR gains produced by processing with the FMV, Frost, and GSC algorithms are 9.14, 5.41, and 4.30 dB, respectively, indicating that the FMV most effectively reduced the interference. This is supported by Figs. 7(c)–(e), which show the spectrograms of the outputs of the FMV, Frost, and GSC algorithms, respectively. It is evident that the FMV output [Fig. 7(c)] more closely resembles the target signal [Fig. 7(a)].

The better performance of the FMV in terms of the SNR gain may be explained by examining the time-varying beam patterns for the FMV, Frost, and GSC algorithms, shown in Figs. 8(a), (b), and (c), respectively. The nulls placed by the FMV converge consistently and quickly to $\pm 45°$. In contrast, the Frost algorithm's nulls rarely converge to exact the directions of the interference sources, and the GSC algorithm's convergence is even less precise. Figures 8(d), (e), and (f) show instantaneous beam patterns for the FMV, Frost, and GSC algorithms. At this point in time, the FMV algorithm placed nulls closer to the directions of the interferers (at $\pm 45°$), especially for the $-45°$ source between 2300 and 5000 Hz.

This example illustrates the typical behavior of the algorithms for signals containing multiple speech sources. When compared to the Frost and GSC algorithms, the FMV algorithm exhibits faster, more accurate adaptation. This results in better interference cancellation.

D. Computational complexity

The computational complexity of the FMV algorithm is derived after some simplifications of the computation. Time-domain microphone data are real-valued, so weights for only $N/2+1$ frequency bins are calculated (where N is the number of FFT bins), and the remaining $N/2-1$ bins are obtained by utilizing the conjugate-symmetry property of the FFT. Because cross-correlation terms in $\mathbf{R}$ are conjugate symmetric, only three of the four terms need to be calculated. For an n-microphone system, weights for only $n-1$ channels

TABLE IV. Computational expense for various parts of the FMV algorithm.

Part of algorithm	Complex adds	Complex multiplies	Complex divisions	Real multiplies	Time
Windowing[a]	0	0	0	$n \cdot N$	$\frac{L}{f_s}$
FFTs	$(n+1)\cdot\frac{N}{2}\log_2 N$	$(n+1)\cdot\frac{N}{4}\log_2 N$	0	0	$\frac{L}{f_s}$
Correlation	$\frac{(n^2+n)\left(\frac{N}{2}+1\right)}{2}$	$\frac{(n^2+n)\left(\frac{N}{2}+1\right)}{2}$	0	0	$\frac{L}{f_s}$
Weight application	$(n-1)\left(\frac{N}{2}+1\right)$	$n\left(\frac{N}{2}+1\right)$	0	0	$\frac{L}{f_s}$
Matrix inversion	$\frac{n^3}{3}$	$\frac{n^3}{3}$	0	0	$\frac{BL}{f_s}$
Weight calculation (not including matrix inversion)	$(n^2+n-2)\left(\frac{N}{2}+1\right)$	$(n^2+n)\left(\frac{N}{2}+1\right)$	$(n-1)\left(\frac{N}{2}+1\right)$	0	$\frac{BL}{f_s}$

[a]For example, the windowing operation requires $n \cdot N$ real multiplies every L/f_s seconds, where L is the number of samples between FFTs, and f_s is the sampling rate in Hz.

need be calculated using Eq. (4), because Eq. (3b) may be utilized to obtain weights for the nth channel with less computation.

The computational complexity of the FMV is a function of f_s, the sampling frequency, n, the number of sensors, B, the block length (the number of FFTs, for a single channel, that are calculated between each computation of new filter weights), N, the number of points in the FFT, and L, the number of output samples obtained from each IFFT. (It is assumed that there is no overlap of groups of output samples.) Assumptions made for this analysis are: (1) a radix-2-real-valued FFT is calculated every L samples; (2) calculating a matrix inverse requires $n^3/3$ complex multiplications and $n^3/3$ complex additions; and (3) the second weight is obtained using Eq. (3b).

Table IV summarizes the costs of each part of the algorithm in terms of complex additions, complex multiplications, complex divisions, real multiplications, and the time in which each must be completed. These cost estimates do not include the cost of overhead within a computer program, such as retrieving data from memory; they reflect only the mathematical operations required to execute the algorithm.

Assuming that two real additions for each complex addition, four real multiplications and two real additions for each complex multiplication, and two complex multiplications and two real divisions for each complex division are required, the total number of operations per second (OPS) required for the algorithm is

$$\mathrm{OPS}=\frac{f_s}{L}\left[N\left(n+\frac{5}{2}(n+1)\log_2 N\right)+(N+2)\left(\frac{4}{3B}n^3+\left(2+\frac{4}{B}\right)n^2+\left(20+\frac{4}{B}\right)n-15-\frac{2}{B}\right)\right]. \tag{12}$$

Additionally, the cost of the Frost and GSC algorithms are, respectively,

$$\mathrm{OPS}_F=f_s\cdot(2K_F+4K_F\cdot n), \tag{13}$$

$$\mathrm{OPS}_{\mathrm{GSC}}=f_s\cdot(n^2+4K_{\mathrm{GSC}}\cdot(n-1)). \tag{14}$$

For the best parameter sets, the FMV and P–K algorithms require 178 million OPS, the Frost algorithm requires 88 million OPS, and the GSC requires 36 million OPS. As implemented, the FMV requires about five times more computation than the GSC and about twice as much as the Frost. By increasing L from 16 to 32 in the FMV and P–K algorithms, the cost is roughly cut in half. In practice, we have found that by increasing the interval between calculation of FFTs, the computational cost of the FMV algorithm can be made approximately equal to that of the GSC or Frost algorithms with only a small decrease in performance.

The P–K algorithm implementation was very similar to that of the FMV, because it also performs FFTs periodically and uses correlation matrices and frequency-domain filtering. Its cost was considered to be comparable to that of the FMV algorithm. Thus, with similar computational cost, the FMV and P–K algorithms provide performance superior to that of the Frost and GSC algorithms in terms of SNR gain. By comparison, the cost of the direct inversion of a time-domain correlation matrix (as per Capon [1969]) of dimension 401 (the length of the Frost and GSC filters), done every 32 samples (as with the calculation of the FMV weights), is approximately 15 billion OPS.

VI. CONCLUSION

Test signals containing multiple speech sources were created from recordings made with three different types of two-microphone arrays in three rooms with varying reverberation times. The performance of a frequency-domain minimum-variance distortionless-response (FMV) beamformer, the Frost adaptive beamformer, the generalized sidelobe canceler (GSC), and an implementation of the Peissig–Kollmeier (P–K) binaural algorithm were evaluated. The FMV and P–K algorithms outperform the time-domain Frost and GSC algorithms in terms of output SNR. A pair of cardioid microphones yields the best performance and lowest target distortion. A pair of omnidirectional microphones performs almost as well, followed by microphones mounted in each of the ear canals of a KEMAR.

The performances of the FMV and P–K algorithms are very similar in terms of the SNR of the output signal, but the P–K algorithm generates significantly more distortion of the target signal. In the tests conducted in this study, filters were used to match approximately the response of the microphones for sounds received from the target direction, but some inherent error is still present in the test signals. This error means that even the distortionless response beamformers could distort the target signal, as is evidenced in the plots of the target distortion. (This issue has been addressed by Elledge [2000].) The FMV generally produces the smallest amount of distortion, while the Frost and P–K algorithms generally produce the largest distortion figures. The impact of the distortion on speech perception and quality has not been determined with human listeners.

Informal observations indicate that the P–K algorithm, as implemented here, produces highly intelligible, though distorted output. This suggests that if distortionless response is a requirement (such as in a high-fidelity hearing aid), then the FMV algorithm is preferred, but if maximal intelligibility is the principal goal, then the P–K algorithm (as implemented here) is also promising. A hybridization of the FMV and P–K algorithms may prove beneficial if it can preserve some of the benefits of each approach.

The two-channel FMV has been implemented in a real-time system [Elledge *et al.*, 2000]. Preliminary tests show that the real-time system helps normal-hearing listeners understand an on-axis talker amidst many interfering sources in controlled environments as compared to conditions with dichotic unprocessed signals and diotic signals (channels summed) [Larsen *et al.*, 2001; Feng, 2002]. Informal tests indicate that the FMV improves intelligibility in rooms that vary widely in size and reverberation characteristics. The output of the real-time system sounds natural and is of high fidelity.

This study uses recordings that were made at a distance of 75 cm from the loudspeaker, where the direct-to-reverberant ratio is relatively high. Even though reverberation was not dominant, the performance of all the algorithms

is compromised to varying degrees, with the FMV and P–K algorithms showing the smallest degradation. Our goal for the present study was to determine the relative effectiveness of these algorithms in rooms having different reverberant characteristics, but not to quantify the effect of the reverberation *per se*. Future work is required to evaluate the performance of the algorithms with microphones farther from the sources where the direct-to-reverberant ratio is lower. Additionally, the 7.6-cm-diameter loudspeakers in the array are quite directional at high frequencies due to the directivity of the driver, thereby reducing the amount of high-frequency reverberation. In the future, loudspeakers with less directional responses should be used in tests that involve sources at greater distances.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Institute on Deafness and Other Communication Disorders (R21DC-04840) of the National Institutes of Health and by grants from the University of Illinois at Urbana-Champaign (the Beckman Institute, the Mary Jane Neer Research Fund, and the Charles M. Goodenberger Fund). We thank Chen Liu and Kong Yang for their earlier work on algorithm simulations, Mark Elledge and Jeffrey Larsen for their participation in the group, and the many staff who generously donated their time to participate in the recording of the speech materials used in our tests.

Berghe, J. V., and Wouters, J. (**1998**). ''An adaptive noise canceller for hearing aids using two nearby microphones,'' J. Acoust. Soc. Am. **103**, 3621–3626.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (**1984**). ''Standardization of a test of speech perception in noise,'' J. Speech Hear. Res. **27**, 32–48.

Brandstein, M., and Ward, D. (**2001**). *Microphone Arrays: Signal Processing Techniques and Applications* (Springer, Berlin).

Capon, J. (**1969**). ''High-resolution frequency-wavenumber spectrum analysis,'' Proc. IEEE **57**(8), 1408–1419.

Cox, H., Zeskind, R. M., and Kooij, T. (**1986**). ''Practical supergain,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-34**(3), 393–398.

Cox, H., Zeskind, R. M., and Owen, M. M. (**1987**). ''Robust adaptive beamforming,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**(10), 1365–1376.

Deslodge, J. G. (**1998**). ''The location-estimating null-steering (LENS) algorithm for adaptive microphone-array processing,'' Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Elledge, M. E. (**2000**). ''Real-time implementation of a frequency-domain array processor,'' M.S. thesis, University of Illinois, Urbana, IL.

Elledge, M. E., Lockwood, M. E., Bilger, R. C., Feng, A. S., Jones, D. L., Lansing, C. R., O'Brien, W. D., and Wheeler, B. C. (**2000**). ''Real-time implementation of a frequency-domain beamformer on the TI C62X EVM,'' presented in 10th Annual DSP Technology Education and Research Conference Houston, TX, 2–4 August.

Feng, A. S. (**2002**). ''Biologically inspired binaural hearing aid algorithms: Design principles and effectiveness,'' J. Acoust. Soc. Am. **111**, 2354.

Frost, O. L. (**1972**). ''An algorithm for linearly constrained adaptive array processing,'' Proc. IEEE **60**(8), 926–935.

Golub, G., and Van Loan, C. (**1996**), *Matrix Computations*, 3rd ed. (The Johns Hopkins University Press Ltd., London).

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. (**1998**). ''Modified LMS algorithms for speech processing with an adaptive noise canceller,'' IEEE Trans. Speech Audio Process. **6**, 338–351.

Greenberg, J. E., Deslodge, J. G., and Zurek, P. M. (**2003**). ''Evaluation of array-processing algorithms for a headband hearing aid,'' J. Acoust. Soc. Am. **113**, 1646–1657.

Griffiths, L. J., and Jim, C. W. (**1982**). ''An alternative approach to linearly constrained adaptive beamforming,'' IEEE Trans. Antennas Propag. **AP-30**(1), 27–34.

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.

Joho, M., and Moschytz, G. S. (**2000**). ''Connecting partitioned frequency-domain filters in parallel or in cascade,'' IEEE Trans. Circuits Syst. **47**, 685–698.

Kates, J. M., and Weiss, M. R. (**1996**). ''A comparison of hearing-aid array-processing techniques,'' J. Acoust. Soc. Am. **99**, 3138–3148.

Kollmeier, B. (**1997**). *Psychoacoustics, Speech and Hearing Aids* (World Scientific, Singapore).

Kollmeier, B., Peissig, J., and Hohmann, V. (**1993**). ''Real-time multiband dynamic compression and noise reduction for binaural hearing aids,'' J. Rehabil. Res. Dev. **30**(1), 82–94.

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.

Larsen, J. B., Lockwood, M. E., Lansing, C. R., Bilger, R. C., Wheeler, B. C., O'Brien, Jr., W. D., Jones, D. L., and Feng, A. S. (**2001**). ''Performance of a frequency-based minimum-variance beamforming algorithm for normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **109**, 2494.

Liu, C., Feng, A. S., Wheeler, B. C., O'Brien, W. D., Bilger, R. C., and Lansing, C. R. (**1997**). ''Speech enhancement via directional hearing based on the place theory,'' 2nd Biennial Hearing Aid Research & Development Conference, Washington, DC, Sept. 1997, 58A (II-P-29).

Liu, C., Wheeler, B. C., O'Brien, Jr., W. D., Bilger, R. C., Lansing, C. R., and Feng, A. S. (**2000**). ''Localization of multiple sound sources with two microphones,'' J. Acoust. Soc. Am. **108**, 1888–1905.

Liu, C., Wheeler, B. C., O'Brien, Jr., W. D., Lansing, C. R., Bilger, R. C., Jones, D. L., and Feng, A. S. (**2001**). ''A two-microphone dual delay-line approach for extraction of a speech sound in the presence of multiple interferers,'' J. Acoust. Soc. Am. **110**, 3218–3231.

Lockwood, M. E. (**1999**). ''Development and testing of a frequency-domain minimum-variance algorithm for use in a binaural hearing aid,'' M.S. thesis, University of Illinois, Urbana, IL.

Lockwood, M. E., Jones, D. L., Bilger, R. C., Elledge, M. E., Feng, A. S., Gouegou, M., Lansing, C. R., Liu, C., O'Brien, Jr., W. D., and Wheeler, B. C. (**1999**). ''A minimum-variance frequency domain algorithm for binaural hearing aid processing,'' J. Acoust. Soc. Am. **106**, 2278.

McDonough, R. N. (**1979**). ''Application of the maximum-likelihood method and the maximum-entropy method to array processing,'' in *Nonlinear Methods of Spectral Analysis: Topics in Applied Physics*, edited by S. Haykin (Springer, New York), pp. 181–243.

Peissig, J., and Kollmeier, B. (**1997**). ''Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners,'' J. Acoust. Soc. Am. **101**, 1660–1670.

Slyh, R. E., and Moses, R. L. (**1993**). ''Microphone array speech enhancement in overdetermined signal scenarios,'' ICASSP-93, **2**, 347–350.

Van Veen, B. D., and Buckley, K. M. (**1988**). ''Beamforming: A versatile approach to spatial filtering,'' IEEE ASSP Magazine, April, 4–24.

Welker, D. P., and Greenberg, J. E. (**1997**). ''Microphone-array hearing aids with binaural output II. A two-microphone adaptive system,'' IEEE Trans. Speech Audio Process. **5**(6), 543–551.

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.

Yang, K. L., Lockwood, M. E., Elledge, M. E. and Jones, D. L. (**2000**). ''A comparison of beamforming algorithms for binaural acoustic processing,'' Proc. 9th IEEE Digital Signal Processing Workshop, Hunt, TX.

Methodology for rheological testing of engineered biomaterials at low audio frequencies

Ingo R. Titze
Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242-1012 and National Center for Voice and Speech, Denver Center for the Performing Arts, Denver, CO 80204

Sarah A. Klemuk
National Center for Voice and Speech, Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242-1012

Steven Gray
Department of Surgery, Otolaryngology, Head-Neck Surgery, Salt Lake City, UT 84113

(Received 20 February 2003; revised 17 October 2003; accepted 20 October 2003)

A commercial rheometer (Bohlin CVO120) was used to mechanically test materials that approximate vocal-fold tissues. Application is to frequencies in the low audio range (20–150 Hz). Because commercial rheometers are not specifically designed for this frequency range, a primary problem is maintaining accuracy up to (and beyond) the mechanical resonance frequency of the rotating shaft assembly. A standard viscoelastic material (NIST SRM 2490) has been used to calibrate the rheometric system for an expanded frequency range. Mathematically predicted response curves are compared to measured response curves, and an error analysis is conducted to determine the accuracy to which the elastic modulus and the shear modulus can be determined in the 20–150-Hz region. Results indicate that the inertia of the rotating assembly and the gap between the plates need to be known (or determined empirically) to a high precision when the measurement frequency exceeds the resonant frequency. In addition, a phase correction is needed to account for the magnetic inertia (inductance) of the drag cup motor. Uncorrected, the measured phase can go below the theoretical limit of $-\pi$. This can produce large errors in the viscous modulus near and above the resonance frequency. With appropriate inertia and phase corrections, $\pm 10\%$ accuracy can be obtained up to twice the resonance frequency.
[DOI: 10.1121/1.1631941]

PACS numbers: 43.80.Cs, 43.80.Ev [FD] Pages: 392–401

I. INTRODUCTION

The larynx and the ear are two organs of the body in which tissues are vibrated in the audio frequency range (20–20 000 Hz). Vibrations in the ear cover the full audio range (by definition), but amplitudes of vibration are relatively small (on the order of 0.01 mm on the eardrum and stapes; Guinan and Peake, 1967). Vibrations in the larynx (i.e., the vocal folds) span a smaller frequency range (50–1000-Hz fundamental frequency), but the amplitudes of vibration are about 100 times larger, on the order of 1 mm (Titze, 1989). The amplitudes depend on the driving pressures and the viscoelastic properties of the tissues. Little scientific effort has been expended to measure these viscoelastic properties in their appropriate frequency ranges. The recent measurements by Chan and Titze (1998; 1999a, b) are a significant beginning, but the reported frequency range of 0–30 Hz is still rather limiting in its application to phonation.

A further limitation is that the majority of rheological measurements has been made on nonviable (dead) tissues. This has been acceptable, for the most part, because the tissue properties of interest are largely determined by the extracellular matrices (Hirano, 1975), which don't change as rapidly after cell death as cellular tissue. But, current interest is in bio-engineered tissues, for which the properties need to be known while the materials are being cultured, i.e., in a viable state. While the focus of this paper is not on measurement of viable tissue *per se*, the methodology for measuring the rheological properties of viable tissue is outlined. The frequency range of interest is 0–150 Hz, which is a modest beginning for fundamental frequencies of phonation.

The specific aims of this report are to describe (1) the optimum gap and plate dimensions for a plate-on-plate rheometer operating under oscillatory shear conditions; (2) the mechanical inertia and phase corrections needed to take the measurements beyond the resonance frequency of the system; and (3) the strain regime needed to match conditions of vocal-fold vibration. The additional methodological problem of how to maintain cell viability during measurement will be touched upon, but details will be left for a subsequent paper.

II. DESCRIPTION OF THE RHEOMETER

The rheometer selected for measurement of shear elasticity and shear viscosity in the 0–150-Hz range is the Bohlin CVO120 high-resolution system (Bohlin Instruments, East Brunswick, NJ). This instrument was selected because of its reported ability to record elastic and viscous moduli at frequencies up to 150 Hz. It is equipped with a drag cup motor that delivers either controlled shear stresses or controlled shear strains to user selected attachments, or measuring systems (Fig. 1). The controlled strain mode is less reliable, however, than the controlled stress mode in this model. Hence, we will address the controlled stress mode only. The

FIG. 1. Rheometer attachments for measuring the viscoelastic properties of a bio-engineered material in a cup-and-plate system filled with culture medium.

measuring systems typically used for biomaterials are flat plates of various radii that make frictional contact with the sample. The sample can be submerged in a cup filled with culture medium as shown in the figure. A peripheral temperature control unit is connected to a water jacket assembly on the base of the rheometer to maintain a set temperature within $\pm 0.05\,^{\circ}\mathrm{C}$.

Measured response to the applied shear stress is transferred from the rheometer firmware to Bohlin software, which displays the results on a computer operating under Microsoft Windows®. The user is able to control the applied torque by setting target shear stresses (or strains) and various other test parameters that are best suited for a given sample. Raw data, as well as reduced data, are accessible in real time and saved in files that can then be accessed by other software for further data analysis.

III. MECHANICAL RESPONSE OF THE SYSTEM

The accuracy of the viscoelastic measurements, as well as the magnitude of the strain that can be delivered to the material, depends on the mechanical response of the rotating motor–shaft–plate assembly (Ferry, 1980; Macosko, 1994). Figure 2(a) shows a sketch of the sample between the plates. The equation of motion for the entire moving assembly can be written as

$$T = I\ddot{\theta} + D\dot{\theta} + K\theta, \tag{1}$$

where T is the torque applied by the motor, I is the overall inertia of the rotating assembly, D is a damping coefficient produced by the sample and (at low frequencies) the drag on the rotating shaft, K is the torsional stiffness of the sample, and θ is the rotational angle (with ''dots over'' indicating time differentiation).

For sinusoidal ($e^{i\omega t}$) motion, the complex response is

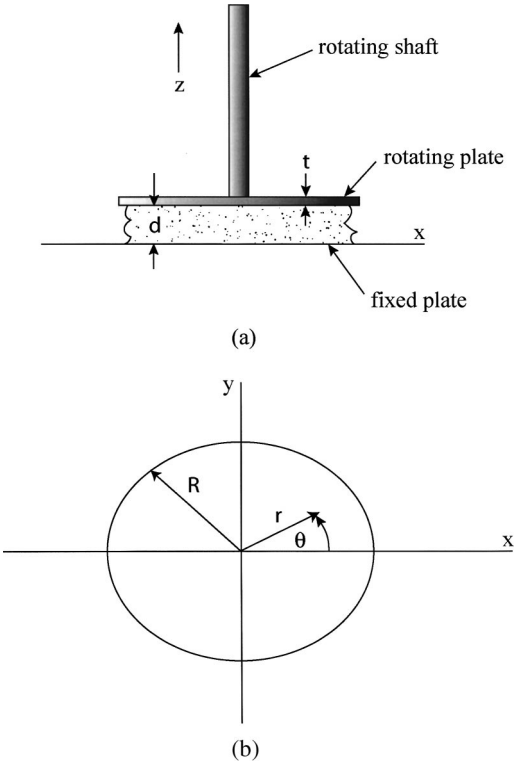


FIG. 2. Sketch of plate-on-plate geometry for determining the mechanical response of the system (a) side view; (b) top view of rotating plate.

$$\frac{\theta^*}{T^*} = \frac{1}{-\omega^2 I + i\omega D + K}, \tag{2}$$

which shows a natural damped resonance frequency at

$$\omega_n = (K/I - D^4/4I^2)^{1/2}. \tag{3}$$

The shear strain γ in the sample between the plates is

$$\gamma = \frac{d\psi}{dz} = \frac{r\theta}{d}, \tag{4}$$

where ψ is the azimuthal (arc) displacement of the moving plate at radius r [Fig. 2(b)], z is the vertical distance into the sample from the bottom plate [Fig. 2(a)], and d is the gap between the plates. The elastic shear stress at any radius r can be written as the product of the elastic shear modulus G' and the shear strain

$$\tau(r) = G'\frac{d\psi}{dz} = G'\frac{r\theta}{d}, \tag{5}$$

which is seen to vary linearly with radius r and rotational angle θ.

The differential restoring torque applied by the sample (due to tissue stiffness) at a radius r is

$$dT_s = r\,dF = r\tau(r)dA = r\tau(r)2\pi r\,dr. \tag{6}$$

Substituting $\tau(r)$ from Eq. (5) into Eq. (6) and integrating over the entire plate

$$T_s = \int_0^R rG'\frac{r\theta}{d}2\pi r\,dr \tag{7}$$

$$= \frac{\pi G' \theta R^4}{2d}, \tag{8}$$

where R is the radius of the moving (upper) plate. Taking the ratio (T_s/θ) yields the shear stiffness of the sample

$$K = (T_s/\theta) = \frac{\pi G' R^4}{2d}. \tag{9}$$

Assuming that the motor–shaft assembly adds or subtracts no rotational stiffness, this stiffness K represents the entire stiffness of the system in Eq. (1). (This assumption does not hold when testing solid materials for which the elastic shear modulus approaches that of the plates.)

In exactly the same manner, the viscous drag torque applied by the sample can be computed. The only difference is that the elastic shear modulus G' is replaced by the shear viscosity η, such that

$$D = \frac{\pi \eta R^4}{2d}. \tag{10}$$

Except at very low frequencies, it can also be assumed that the motor–shaft assembly has negligible friction (by being built with a forced-air suspension system that drastically reduces surface contact). Then, substitution of Eqs. (9) and (10) into Eq. (2) yields the frequency response

$$\frac{\theta^*}{T^*} = \frac{2d}{G' \pi R^4} \frac{1}{1-(\omega/\omega_0)^2 + i\omega\eta/G'}, \tag{11}$$

where θ^* is the complex angular displacement, T^* is the complex torque, and ω_0 is the undamped resonance frequency, obtained from Eq. (3) by setting $D=0$

$$\omega_0 = (K/I)^{1/2} = \left(\frac{\pi G' R^4}{2dI}\right)^{1/2}. \tag{12}$$

The damped resonance frequency from Eq. (3) becomes

$$\omega_n = R^2 \left(\frac{\pi G'}{2dI} - \pi^2 R^2 \eta^2/(16 d^2 I^2)\right)^{1/2} \tag{13}$$

Recall from Eq. (4) that the shear strain varies linearly with the radial dimension. In the Bohlin CVO 120 rheometer, a *representative* shear strain is chosen to be at 75% of the radius of the plate, such that

$$\gamma^* = 0.75 \frac{R\theta^*}{d}. \tag{14}$$

With this choice, the ring outside this $3/4R$ radius (56.25% of the sample area) experiences greater than this representative strain, and the inner circle (43.75% of the area) experiences less than this representative strain. The complex shear response now becomes

$$\frac{\gamma^*}{T^*} = \frac{1.5}{G' \pi R^3} \frac{1}{1-(\omega/\omega_0)^2 + iG''/G'} = Ae^{i\phi}, \tag{15}$$

where $G''=\omega\eta$, A is the magnitude of the response, and ϕ is the phase. For $\omega \to 0$, the static strain magnitude is $1.5/(G'\pi R^3)$ per unit torque $(N-m)$, and the phase ϕ is $-\tan^{-1} G''/G'$. The strain at the undamped resonance frequency ($\omega=\omega_0$) is a factor G'/G'' larger than at $\omega=0$, with a $-90°$ phase. The ratio G''/G' is known as the loss tangent of the material, or its damping ratio.

IV. SENSITIVITY OF RESPONSE TO SYSTEM PARAMETERS

Figure 3 shows families of theoretical resonance curves (magnitude of $\gamma^*/T^*=A$) for variable plate radius R, variable gap size d, variable elastic shear modulus G', and variable loss tangent G''/G'. The motor–shaft inertia I was set to be $1.89\times10^{-5}\,\mathrm{kg\,m^2}$ (see the detailed description later). Whenever the variables were held constant, they were at the nominal (midrange) values $R=20\,\mathrm{mm}$, $d=1.2\,\mathrm{mm}$, $G'=4.0\,\mathrm{kPa}$, and $G''/G'=0.6$. Note that the resonance frequency increases with increasing plate radius R (upper left), but the magnitude of the response decreases sharply with increasing R. Thus, some compromise must be reached between the amount of shear strain that can be applied and the frequency of oscillation. Reducing the gap size does not affect the response magnitude, but raises the resonance frequency (upper right). Gap size is the most important variable for audio frequency applications. It enters Eq. (15) through ω_0 [recall Eq. (12)].

Now, consider the material properties themselves as rheometer variables. Increased elastic shear modulus G' of the sample increases the resonance frequency, but decreases the strain magnitude for constant torque (lower left). Loss tangent, as expected, has an effect only on the response magnitude (lower right), but not the resonance frequency. Lower loss tangents produce larger strains for the same amount of applied torque.

The resonance curves allow the sensitivity of the rheometer to be ascertained for changing viscoelastic constants. On the left side of the resonance frequency in each curve of Fig. 3, the mechanical impedance is ''elastic;'' i.e., it is under control of the elastic properties of the tissue sample. In this low-frequency region, the elastic shear modulus G' has a large effect on the response magnitude (lower left figure). Because the curves are spread far apart, measurements of G' would appear to have their greatest accuracy in this 0–50-Hz region. On the right side of the resonance frequency, however, the response is controlled by the inertia of the motor–shaft assembly. Note that the shear modulus and loss tangent curves all converge, suggesting that the response is insensitive to different values of G'. In this ''inertia-controlled'' region, measurements would appear to be inaccurate unless the system inertia is known to a highly significant figure.

The greatest sensitivity to G''/G', the loss tangent, is near resonance (bottom right figure), but this sensitivity decreases dramatically above resonance (about 35 Hz in this case). Differences in loss tangents between 0.2 and 1.0 are barely discernible at 50 Hz. For this reason, the high-frequency response magnitude can be used to determine the system inertia, as will be discussed later.

The viscoelastic parameters are extracted from measurements of the amplitude and phase response to an applied sinusoidal torque. Substituting ω_0 from Eq. (12) into Eq. (15) and solving for G' and G'', we obtain

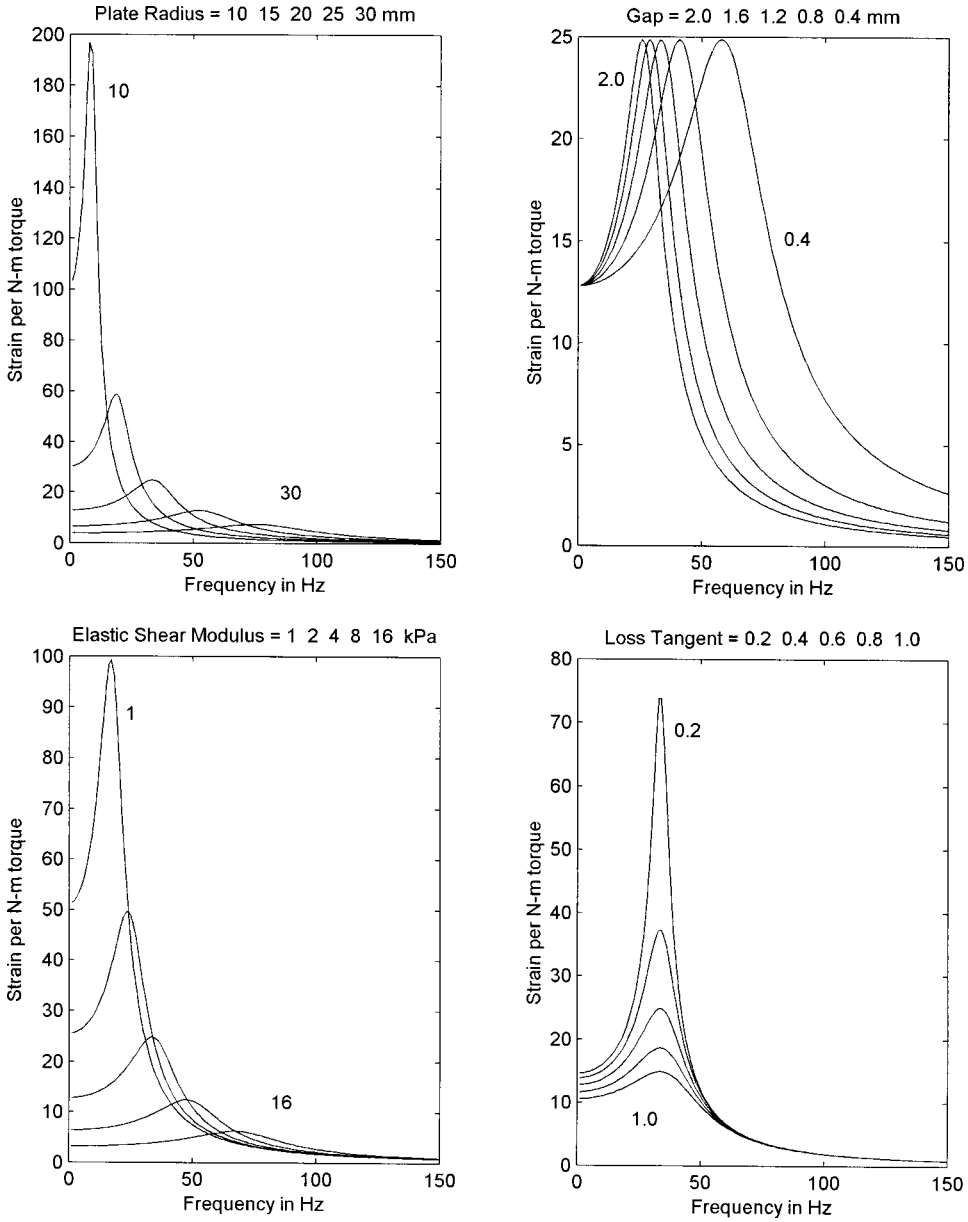


FIG. 3. Variation of strain response with plate radius (upper left), gap size (upper right), elastic shear modulus (lower left), and loss tangent (lower right).

$$G' = \frac{1.5}{\pi R^3}\left[\text{Real}\left(\frac{T^*}{\gamma^*}\right) + \frac{d}{0.75R}\omega^2 I\right]$$

$$= \frac{C_1}{A}\cos\phi + C_1 dI\omega^2/C_2, \tag{16}$$

$$G'' = \frac{1.5}{\pi R^3}\text{Imag}\left(\frac{T^*}{\gamma^*}\right) = -\frac{C_1}{A}\sin\phi, \tag{17}$$

where $C_1 = 1.5/(\pi R^3)$ and $C_2 = 0.75R$. Note that the accuracy of the extraction of G' depends on the relative size of two terms on the right in Eq. (16). If the second term is much larger than the first (usually for large ω), the extraction becomes inaccurate unless the values of I and d, are known precisely. This is shown in Fig. 4, where a $\pm 10\%$ error has been introduced into the inertia I, which was nominally $1.89\times 10^{-5}\,\text{kg m}^2$. This error (dashed lines) appears relatively small in the magnitude and phase response curves (upper plots). However, the extraction of the elastic modulus with Eq. (16) shows a huge error range (lower left plot, dashed lines). For low frequencies, the extracted value of G' is 4.0 kPa, the theoretical value. The error increases quadratically with frequency, however, reaching $\pm 100\%$ at 100 Hz. Improvements in accuracy can be made by raising the resonance frequency, i.e., by reducing the gap size, reducing the inertia, or increasing the plate radius, all of which make the second term in Eq. (16) as small as possible.

The viscous modulus G'' is unaffected by I and d, as shown in Eq. (17) and the lower right panel of Fig. 4. However, since G'' is determined by both magnitude and phase, and since the magnitude is inversely proportional to G' [recall Eq. (15)], any error in G' will be transferred to an error in G'' if the magnitude measurement is prone to error. Fortunately, the magnitude measurement is relatively accurate in most stress-controlled rheometers. However, the phase response is quite error prone.

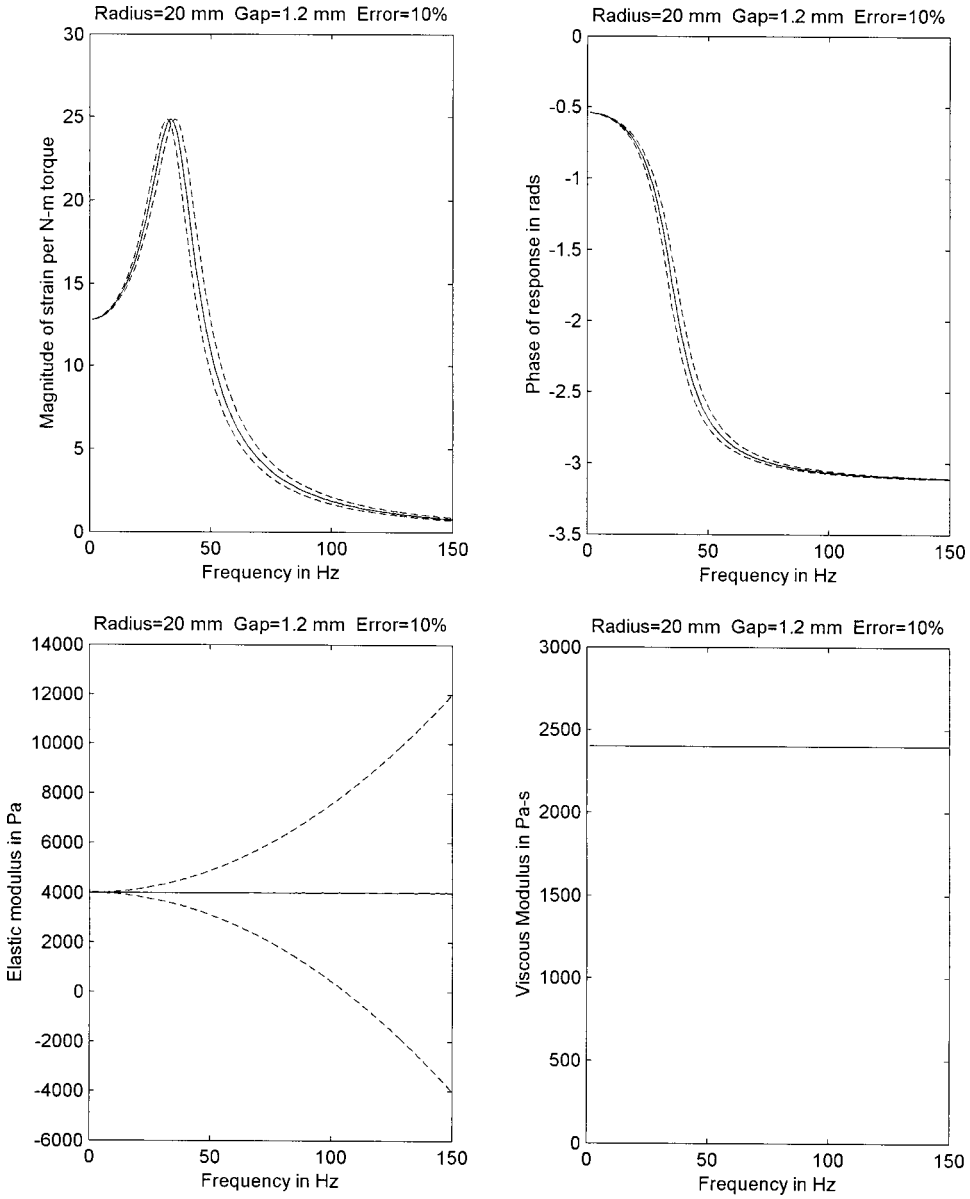


FIG. 4. Magnitude and phase response of rheometer (above) and extracted G' and G'' when a $\pm 10\%$ error is introduced into the system inertia (dashed lines).

V. PHASE CORRECTIONS

A viscoelastic material with known properties (NIST SRM 2490 polyisobutylene dissolved in 2,6,10,14-tetramethylpentadecane) was chosen as a test material (Schultheisz and Leigh, 2002). It is a non-Newtonian standard that exhibits substantial shear thinning and significant elasticity. The manufacturer provided tabular data from $\omega=0.04$ rad to $\omega=100$ rad (0.0064 to 15.9 Hz), which unfortunately did not reach the audio range. Figure 5 shows the frequency variation of G' and G'' as specified by the manufacturer for 20 °C (asterisks). A fifth-order polynomial least-squares fit was used to extrapolate the data to 150 Hz using the MATLAB *polyfit* routine (solid lines). The extrapolation was performed in log–log space to make use of three decades of frequency measurements with equal spacing. Superimposed on Fig. 5 are additional measurements in the 1–40-Hz range performed in our laboratory on the same material at the same temperature (small dots). This range partially overlaps with the manufacturer's data (from 1–16 Hz) but does not venture beyond the resonance frequency (approximately 40 Hz, to be discussed below). A second-order polynomial extrapolation was used in this case (again using MATLAB *polyfit*) to extend the 1 to 40-Hz range of measurements to 150 Hz (dashed lines). The second-order extrapolation (rather than a fifth order) was used because there were fewer points to work with. It is evident from the figure that differences between extrapolations can occur, depending on the order of the extrapolation, the number of data points used, and the range of the extrapolation. The solid curves of Fig. 5 will be used later as a ''standard'' of comparison with measured values, but given that true measurements were not available at the higher frequencies, it is considered only a pseudostandard.

Figure 6 shows the response curves calculated from Eq. (15) with measured and extrapolated values of G' and G''. As before, the Bohlin CVO120 data are shown with dots and the theoretical extrapolations are shown with solid and dashed lines. The upper left is the real part of (γ^*/T^*), the upper right is the imaginary part, the lower left is the magnitude A, and the lower right is the phase ϕ. Note that the

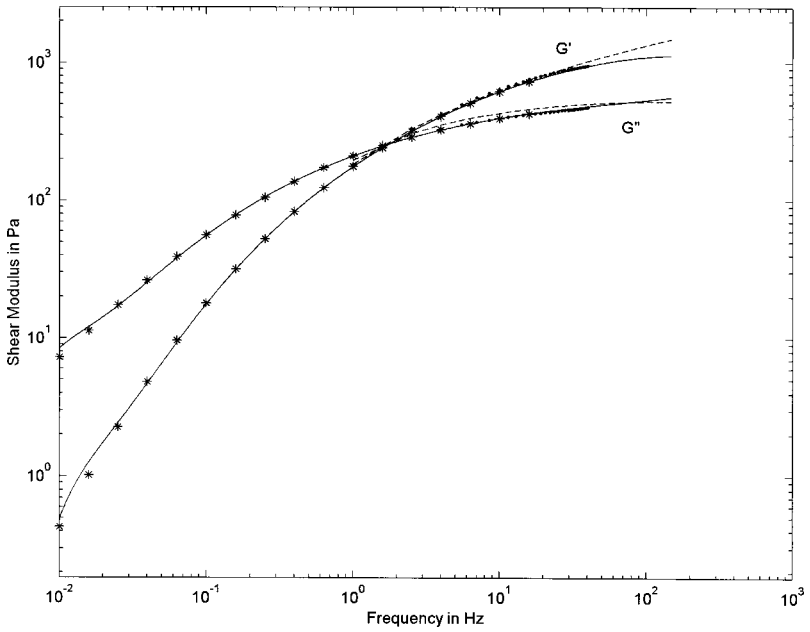


FIG. 5. Elastic modulus G' and viscous modulus G'' as reported by the manufacturer for a calibration material at 20 °C. Solid lines are extrapolations.

theoretical and measured magnitudes agree well over the entire range of frequencies (up to 150 Hz). The peak occurs at 38 Hz, which defines the resonance frequency. The measured phase does not agree with theory, however. It drops below the theoretically predicted curve, which accounts for the errors in the real and imaginary parts of the frequency response (top curves). Two phase corrections are also shown (lower right). The dashed curve represents a linear phase correction, derived by assuming that

$$\phi(\omega) = \phi_r(\omega) + \alpha\omega, \tag{18}$$

where $\phi_r(\omega)$ is the raw phase and α is a constant determined as follows. We assume that the phase can be anchored to the theoretical value at a value x times the resonant frequency (typically x is chosen between 2.0 and 3.0), at which point the theoretical phase is

$$\hat{\phi}(x\omega_0) = \tan^{-1}\frac{G''/G'}{|1-x^2|} - \pi, \tag{19}$$

according to Eq. (15). We also assume that the loss tangent does not change much from ω_0 to $x\omega_0$, which means that G''/G' can be evaluated at or near resonance, where measurements are relatively accurate. With these assumptions, the slope α becomes

$$\alpha = \frac{\hat{\phi}(x\omega_0) - \phi_r(x\omega_0)}{x\omega_0}, \tag{20}$$

where $\phi_r(x\omega_0)$ is the measured raw phase at the anchor point. Physically, this linear phase correction is attributed to an inductive reactance of the drag cup motor, which delays the torque in relation to the applied voltage to the motor windings. Note that in Fig. 6 the linearly corrected phase (dashed line) is nearly indistinguishable from the theoretical phase up to about 100 Hz, more than twice the resonance frequency. The value of x was 2.5 and the corresponding α was 0.000 764 s (0.0438 deg/Hz).

A second-order phase correction can be made to the high-frequency tail if measurements are taken beyond $2\omega_0$. From Eq. (15), it is clear that the theoretical phase at any high-frequency ω_h is

$$\hat{\phi}_h = -\text{angle}[1-(\omega_h/\omega_0)^2 + iG''/G']. \tag{21}$$

If we assume that the G''/G' variation is small in relation to $(\omega_h/\omega_0)^2$, and that the value of the loss tangent can once again be taken to be the resonance value, a second-order phase correction becomes

$$\phi_c = \frac{\hat{\phi}_h - \phi_h}{(\omega_h - 2\omega_0)^2}(\omega - 2\omega_0)^2 \quad \omega > 2\omega_0, \tag{22}$$

where ϕ_h is the phase computed from Eq. (18) at ω_h. Equation (22) can then be added to Eq. (16) for frequencies above $2\omega_0$. The dot-dashed line in Fig. 6 (barely visible) below the solid line is the second-order phase correction, which is very close to the theoretical prediction from the calibration data. The value of β was 6.56×10^{-7} s^2/rad.

VI. ACCURATE DETERMINATION OF THE SYSTEM INERTIA

As discussed earlier, inertial drag of the system affects the measurement of G', especially at frequencies above mechanical resonance. The largest component of the inertia is that of the rotor of the drag cup motor, making minimization and accurate knowledge of its value critical for frequencies in the audio range. Since a direct measure of the motor inertia is not feasible (because it involves variable appliances), an experimental value is obtained. As will be seen, however, this experimental value of inertia always involves errors in the gap size d, because mathematically d and I always appear as a product. Hence, the value of I fluctuates more than expected if d is assumed error free.

Newtonian oils that meet National Institute of Standards and Testing (NIST) specifications can be used to get a first estimate of the motor inertia. Since the shear deformation

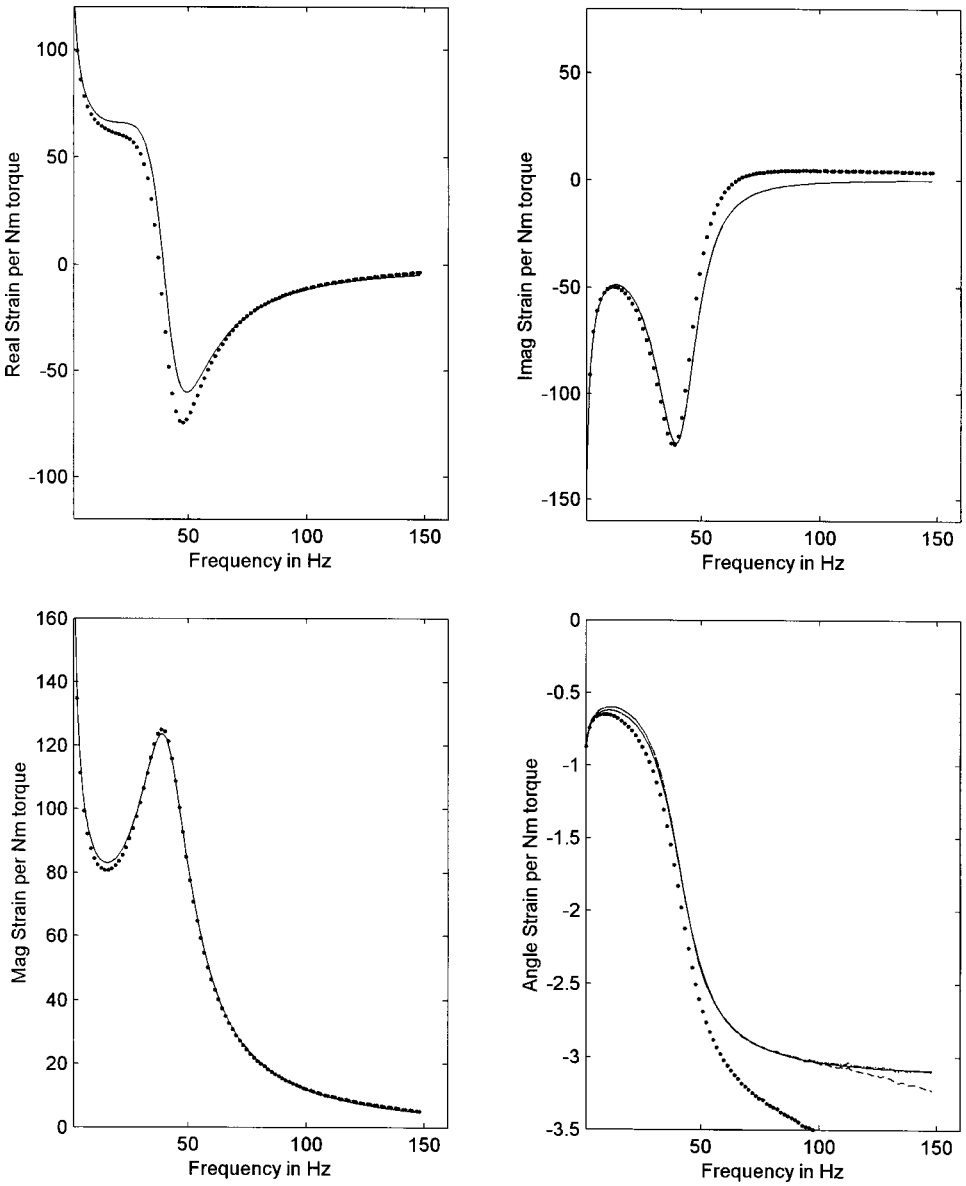


FIG. 6. Measured and corrected values of the strain–torque response. Upper left is the real part, upper right the imaginary part, lower left the magnitude, the lower right the phase. Dots are measured values from the Bohlin rheometer and solid lines are extrapolations of the manufacturer's data.

rate of a purely viscous (Newtonian) material is $-90°$ out of phase with a constant (nonoscillatory) shear torque, one can expect to see this constant $-90°$ phase measurement for all frequencies if the correct value of motor inertia is programed in the analysis software. This is the technique recommended by most rheometer manufacturers, including Bohlin. A check can also be made on dynamic (oscillatory) viscosity with the use of the oil. NIST certifies the dynamic viscosity of its oils to be within a specified tolerance of $\pm 1\%$. During a frequency sweep experiment, the motor inertia can be adjusted to produce dynamic viscosity readings within this margin of error. Unfortunately, this is somewhat of a trial and error process. For the measurements G' and G'' reported above, this method yielded a value of $1.87\times 10^{-5}\,\mathrm{kg\,m^2}$.

The validity of this first estimate is tested against the measured resonance frequency of the system. From Eq. (12), it is evident that the inertia I can be estimated as

$$I=\frac{\pi G_0' R^4}{2d\omega_0^2}, \tag{23}$$

where G_0' is the elastic shear modulus at resonance, which is obtained from the raw data. This calculation resulted in a value of $2.20\times 10^{-5}\,\mathrm{kg\,m^2}$, which is considerably higher than that measured with the Newtonian oil. A discrepancy therefore clearly exists, which can be attributed to either an incorrect G_0' or an incorrect d, if in fact the first estimate I is correct.

A third estimate of I is obtained from the highest frequency of measurement. Recall from Fig. 3 (bottom curves) that the magnitude A of the response is not sensitive to G' or G'' at high frequencies. The denominator of Eq. (15) is dominated by the $(\omega/\omega_0)^2$ term. Considering this term to be the only one of significance, and using Eq. (12) to substitute for ω_0, the inertia is estimated as

$$I=\frac{1.5R}{(\omega_h^2 2d)A_h}, \tag{24}$$

where ω_h is the highest frequency of measurement (usually $>3\omega_0$) and A_h is corresponding response magnitude. This estimate yielded $I=1.69\times 10^{-5}\,\mathrm{kg\,m^2}$, which is lower than

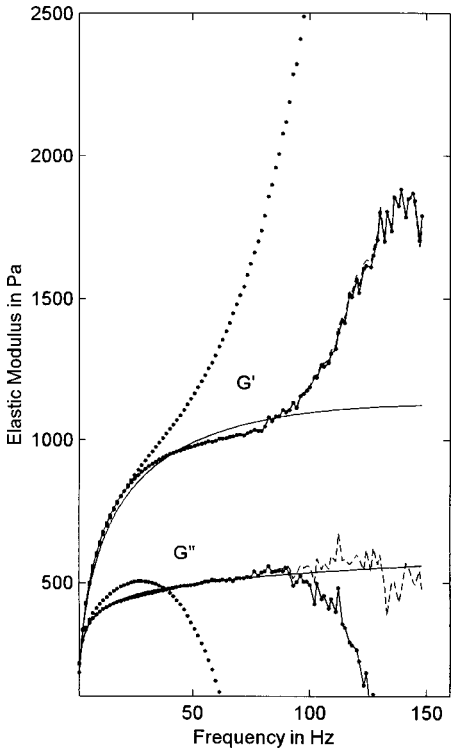


FIG. 7. Measured and corrected values of G' and G'' for the calibration fluid. Unconnected dots are now measured data from the Bohlin rheometer. Connected dots are for a linear phase correction, dashed lines for an additional second-order phase correction, and solid lines are extrapolations of the manufacturer's data.

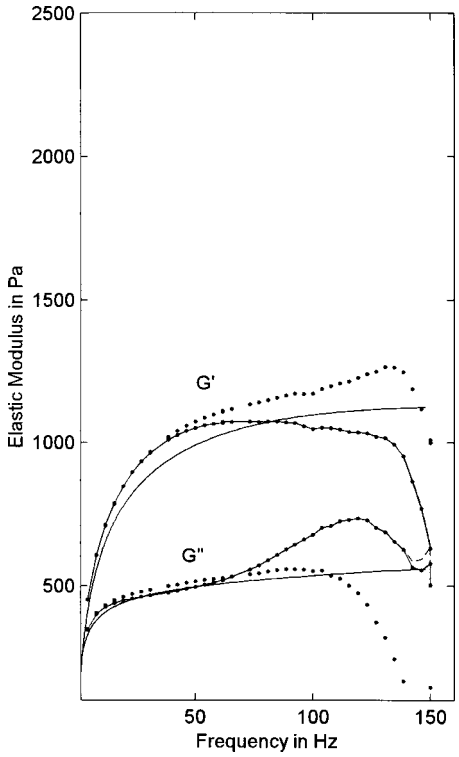


FIG. 8. Same as Fig. 7, but with gap size reduced from 0.2 to 0.1 mm.

both of the above estimates. Taking the average of the values obtained from Eqs. (23) and (24) resulted in an inertia of $1.94\times10^{-5}\,\mathrm{kg\,m^2}$. This value was used in the final computations of G' and G''. Ideally, however, an iterative procedure would be used in which new test runs would be conducted with ever better estimate of I until all three values converge. In many applications, this is impractical because of the time involved.

VII. FINAL RESULTS

Figure 7 shows details of G' and G'' measurements with and without phase corrections. Unconnected dots are for raw data, connected dots for the linear phase correction, and dashed lines for the additional second-order phase correction. The solid lines are the extrapolations of data provided by the manufacturer. Above resonance (38 Hz), the uncorrected values of G' diverge upward and those of G'' diverge downward, suggesting a large error in the phase measurement. [Recall that the sine and cosine in Eqs. (16) and (17) drift in opposite directions.] The downward deviation of G'' is so large that values become zero at about 63 Hz. With the first phase correction, the data are within $\pm5\%$ of the expected values from the manufacturer (up to twice the resonance frequency). Beyond 80 Hz, the second phase correction does improve G'', but has no effect on G'. Noise in the data set reflects uncertainty in the amplitude at high frequencies.

Figure 8 shows data for the same calibration material, but this time the resonant frequency was raised to 63 Hz by reducing the gap size d from 0.2 to 0.1 mm. Similar phase corrections were made and the system inertia was determined by the method described above. The data now appear to have some validity up to 100 Hz, but the phase correction yielded a bump in G'' around 120 Hz because the highest frequency of measurement was only $2.2\omega_0$. This forced the anchor value x to be 2.2 and left few data points for the second-order phase correction.

VIII. ACHIEVEMENT OF STRAINS TYPICAL FOR VOCAL-FOLD VIBRATION

The Bohlin CVO120 rheometer provides a maximum torque of 0.12 N m. The maximum strain achievable at any frequency is this number multiplied by the magnitude A of the strain–torque response (lower left in Fig. 6). For example, at frequencies near zero, a strain of about 14 (1400%) is achievable for this calibration material with a plate radius of 20 mm and a gap size of 0.2 mm. At the resonance frequency, a similar strain is achievable. But, at 150 Hz, the maximum achievable strain has dropped to about 0.6 (5 per N m times 0.12 N m). This strain is still about the maximum shear strain encountered in vocal-fold vibration (Titze, Švec, and Popolo, 2003), suggesting that the targets for simulating vocal-fold vibration are achievable at all frequencies of interest. It must be remembered, however, that the response magnitude reduces sharply with larger elastic and viscous moduli, smaller plate radii, and larger gap sizes. Not all vocal-fold tissues (real or synthetic) spread out nicely over the plates to accommodate the small 0.2-mm gap size, and in some cases the amount of material available is insufficient to

cover the 20-mm plates. In these cases, strains on the order of 0.1–1.0 may be difficult to achieve, except near resonance.

IX. MAINTAINING VIABILITY OF THE CELL CULTURES

In order to simulate vocal-fold vibration on engineered biomaterials as proposed here, homeostatic conditions of the cell culture environment in the rheometer were evaluated. During vibration experiments, the cell medium was subjected to ambient air for periods of 6 to 12 h at a controlled temperature of 37 °C. Parameters considered in this evaluation were *p*H and evaporation of the medium.

Typically, fibroblast cells grown *in vitro* divide and synthesize extracellular material at a *p*H of 7.4–7.7 (Freshney, 2000). This range of *p*H is an indicator of equilibrium between dissolved carbon dioxide (CO_2) and hydrogen carbonate (HCO_3^-) in the liquid medium. Carbonic acid (H_2CO_3) is produced as a result of CO_2 from the ambient air being in contact with the medium. It then dissociates in two directions

$$H_2O + CO_2 \Leftrightarrow H_2CO_3 \Leftrightarrow H^+ + HCO_3^- , \tag{25}$$

with a natural tendency to return to the acidic H_2CO_3 and raising the *p*H. By increasing atmospheric CO_2 and controlling the bicarbonate concentration in the medium, an equilibrium of neutral *p*H can be maintained (Freshney, 2000). For laryngeal fibroblast cultures, this neutral *p*H results from maintaining 5% atmospheric CO_2 in an incubator set at 37 °C and formulating a medium ''cocktail'' that contains the proper balance of salts and amino acids. For example, in one series of measurements, the initial *p*H of the medium was 7.0–7.2. Within 18 min of exposure to room air, the *p*H rose to 7.97, within 45 min to 8.12, and after 8 h to 8.85. At the 8-h marker, nearly all laryngeal fibroblasts were dead, as observed with a 20× light microscope. Attempts to duplicate an exchange of CO_2-rich, humidified air across a confined air/medium interface in the rheometer setup resulted in the same acidic rise and did not deter water evaporation. However, when 1.5 mM of HEPES buffer was included in the medium, *p*H readings were 7.71 after 2.5 h of ambient air exposure, 8.01 after 12 h of exposure, and the laryngeal fibroblasts continued to exhibit viable behavior.

Measured evaporation of aqueous medium indicated significant increases in salt concentration and therefore osmolality. Volume of medium was measured every 45 min for 8 h under rheometer experimental conditions. The mean evaporation rate was 557 μL/h, which meant that more than half of the water content of the medium was gone after 8 h. The effect most certainly contributed to the observed cell death. In order to compensate for this deterioration, it was concluded that a complete exchange of medium should occur every 2 hours. An in-depth treatment of cell culture viability will be addressed in a subsequent paper.

X. CONCLUSIONS

The methodology for measurement of viscoelastic properties of biomaterials with a parallel plate stress-controlled rheometer has been reviewed. Extension of this methodology into the low audio frequency range (50–150 Hz) requires considerable attention to be paid to sample geometry, phase corrections, and determination of system inertia. In general, the plate radius needs to be as large as possible and gap size between the plates as small as possible to set the mechanical resonance frequency high. This requires thin sheets of material to be engineered, which are cut into circular wafers (>20-mm radius and <1.0-mm thickness). These dimensions allow the mechanical resonance frequency to be taken into the 50–150-Hz region, partially avoiding the difficult inertia-controlled region of the rheometer where measurement accuracy is very poor. For a given error in the inertia, the measurement error in G' increases as $(f-f_0)^2$, where f is the measurement frequency and f_0 is the mechanical resonance frequency. In a controlled strain rheometer, it is likely that some of these mechanical resonance problems are canceled out by a control loop, but it is likely that the internal feedback system will make it more difficult for the user to make corrections. We have not dealt with strain-controlled devices, and therefore withhold further comment.

For fluid-like samples with relatively small molecules, small gaps and large radii are not a problem, unless adequate amounts of the fluid are not available to cover the large plates. For materials with large molecules, or engineered tissue scaffolds with large porosity, gap sizes on the order of 0.1 mm may be too small to maintain tissue homogeneity.

If the resonance frequency of the system cannot be set high enough, an extrapolation or phase correction scheme can be used to extend measurements to about twice the resonance frequency. But, phase corrections require preliminary passes through the data to obtain an accurate value of the resonant frequency and the high-frequency response magnitude, and therewith the system inertia.

For tissue viability, the typical plate-on-plate system can be replaced with a plate-in-cup system that allows the engineered tissue to be immersed in a medium. It has been shown that exchange of the medium about once every 2 hours maintains the proper conditions for cells to maintain their viability. A detailed study of viability is underway.

Future methodological investigations will be directed toward adherence of the tissue sample to the plates at low audio frequencies. Molecular adhesion between the plate and the material plays an important role. A certain amount of friction between the sample and the plates can generally be maintained. This is accomplished by either coating the plates with fine-grit nontoxic sandpaper, by roughing the surfaces of the plates, or by applying a normal force that increases the frictional force. However, the normal force may alter the nature of the engineered tissue if too much volumetric compression occurs. We generally have limited the volumetric compression to less than 10%. Slippage does regularly occur for some materials, which needs to be understood and corrected before meaningful data sets can be reported. Such slippage will make the effective gap size different from the true gap size. Since gap size and inertia always appear as a product in the equations of analysis, errors in gap size can be interpreted as errors in inertia. A follow-up article on this topic is in preparation.

ACKNOWLEDGMENT

This work is supported by Grant No. 1 R01 DC04336-02 from the National Institute on Deafness and Other Communication Disorders.

Chan, R., and Titze, I. (**1998**). ''Viscosities of implantable biomaterials in vocal fold augmentation surgery,'' Laryngoscope **108**, 725–731.

Chan, R., and Titze, I. (**1999a**). ''Viscoelastic shear properties of human vocal fold mucosa: Measurement methodology and empirical results,'' J. Acoust. Soc. Am. **106**(4), 2008–2012.

Chan, R., and Titze, I. (**1999b**). ''Hyaluronic acid (with fibronectin) as a bioimplant for the vocal fold mucosa,'' Laryngoscope **109**, 1142–1149.

Ferry, J. D. (**1980**). *Viscoelastic Properties of Polymers*, 3rd ed. (Wiley and Jones, New York).

Freshney, R. I. (**2000**). *Culture of Animal Cells. A Manual of Basic Technique*, 4th ed. (Wiley, New York).

Guinan, J. Jr., and Peake, W. (**1967**). ''Middle-ear characteristics of anesthestized cats,'' J. Acoust. Soc. Am. **41**(5), 1237–1261.

Hirano, M. (**1975**). Phonosurgery: Basic and Clinical Investigations. Kurume, Japan, unpublished official report.

Macosko, C. W. (**1994**). *Rheology: Principles, Measurements, and Applications* (Wiley-VCH, New York).

Schultheisz, C. R., and Leigh, S. F. (**2002**). ''Certification of the rheological behavior of SRM 2490, Polyisobutylene dissolved in 2,6,10,14-tetramethylpentadecane,'' NIST Special Publication 260-143, U.S. Department of Commerce, National Institute of Standards and Technology.

Titze, I. R. (**1989**). ''On the relation between subglottal pressure and fundamental frequency in phonation,'' J. Acoust. Soc. Am. **85**(2), 901–906.

Titze, I. R., Svec, J. G., and Popolo, P. S. (**2003**). ''Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues,'' J. Speech Lang. Hear. Res. **46**, 922–935.

Reverberation and frequency attenuation in forests—implications for acoustic communication in animals

Mark Padgham
Department of Zoology, The University of Melbourne, Australia 3010

(Received 19 July 2002; revised 12 September 2003; accepted 3 October 2003)

Rates of reverberative decay and frequency attenuation are measured within two Australian forests. In particular, their dependence on the distance between a source and receiver, and the relative heights of both, is examined. Distance is always the most influential of these factors. The structurally denser of the forests exhibits much slower reverberative decay, although the frequency dependence of reverberation is qualitatively similar in the two forests. There exists a central range of frequencies between 1 and 3 kHz within which reverberation varies relatively little with distance. Attenuation is much greater within the structurally denser forest, and in both forests it generally increases with increasing frequency and distance, although patterns of variation differ between the two forests. Increasing the source height generally reduces reverberation, while increasing the receiver height generally reduces attenuation. These findings have considerable implications for acoustic communication between inhabitants of these forests, particularly for the perching behaviors of birds. Furthermore, this work indicates the ease with which the general acoustic properties of forests can be measured and compared.
[DOI: 10.1121/1.1629304]

PACS numbers: 43.80.Ev, 43.80.Lb [WA] Pages: 402–410

I. INTRODUCTION

The influences of habitat on the evolution of acoustic communication by animals, particularly birds, were first given serious consideration in the 1970s (Chappius, 1971; Morton, 1975), with broadcast and re-recording experiments (e.g., Dabelsteen *et al.* 1993; Holland *et al.* 1998; Mathevon *et al.*, 1996) often used to examine such influences. Conclusions from these experiments pertain to the particular signals and habitats studied, and commonly do not address the *general relationships* between habitat and patterns of acoustic communication (a notable exception being Wiley and Richards, 1982), even though the importance to animal communication of both attenuation (Wiley and Richards, 1982) and reverberation (Richards and Wiley, 1980) are well established. Some avian species can both assess and respond to reverberation (Naguib, 1995; Wiley and Godard, 1996), and studies of song structures from larger groups of species suggest that reduction of the deleterious effects of reverberation has been an influential factor in the evolution of birdsong (Badyaev and Leaf, 1997; Wiley, 1991).

The present work aims to quantify the nature of reverberative decay within forests, as well as the frequency-attenuating effects of acoustic propagation. Theoretical knowledge of the processes governing reverberation within complex environments such as forests remains limited (cf. Polack, 1992). There have been a number of studies of the acoustic characteristics of vegetation (Aylor, 1972; Martens, 1980; Martens and Michelsen, 1981; Price *et al.*, 1988), but since the pioneering work of Kuttruff (1967), there seem to have been few attempts to measure reverberation within forests and to relate such measurements to theoretical predictions. One notable exception was in the work of Huisman and Attenborough (1991), who found acceptable agreement between measurement and theory, but who also suggested that their theoretical treatment of reverberation was inadequate.

Much acoustic theory proceeds from an assumption of exponential reverberative decay, as exists in the presence of a diffuse acoustic field (e.g., Kuttruff, 2000; Polack, 1992). However, any acoustic field within a forest is not bounded, and thus may not theoretically be presumed to be diffuse. A diffuse acoustic field may nevertheless exist within forests, especially because most of the reflective surfaces within it will almost certainly be randomly arranged and oriented. Establishing the presence of exponential reverberative decay within a forest would greatly aid theoretical treatment of the effects of acoustic propagation, which in turn would facilitate understanding of the evolutionary influences of the acoustic environment upon communication between animals.

For the present work, experiments were conducted within two types of southeast Australian forest, examining relationships between the relative positions of sources and receivers and both rates of reverberative decay and patterns of frequency attenuation. The technique involves recording the impulsive sounds (cf. Schroeder, 1965) of balloon bursts at different distances and heights, a method which has proved convenient to use, can be completed within a day, and is shown to produce highly repeatable measurements. Analyses of balloon-burst recordings furnish data on reverberative decay as well as on frequency attenuation, thus permitting quantification of the major effects on sound of propagation through the forests. Finally, the methods of analysis used yield a fruitful way of comparing the acoustic properties of different environments, which can provide a more comprehensive basis from which to discuss observed differences in acoustic signals used by animals inhabiting different environments. Variation measured with regard to specific signals (e.g., Naguib, 2003; Holland *et al.*, 1998) may be more

TABLE I. General characteristics of the forests within which the experiments were conducted.

Forest	Annual rainfall (mm)[a]	Relative humidity[a,b]	Tree heights (m)	Lower limit of foliage (m)	Typical leaf dimensions (cm)	Under-storey
Box-ironbark[c]	560	48%	10–20	5–10	8–14×1–3	Very little
Mountain ash[d]	1020	73%	45–50	20–25	10–13×3–5	Copious and dense[e]

[a]Figures provided by the Australian Bureau of Meteorology.
[b]Average figures at 3 pm during the months of the experiments.
[c]Dominant species of tree are gray box (*Eucalyptus microcarpa*), red ironbark (*E. sideroxylon*), and yellow gum (*E. leucoxylon*).
[d]The predominant tree species is mountain ash (*E. regnans*), with a few mountain gray gum (*E. cypellocarpa*), and shining gum (*E. nitens*).
[e]Including many trees and shrubs of varying heights, with ground covered by ferns and bracken.

readily understood by reference to knowledge of the general acoustic properties of particular forests.

II. METHODS

The acoustic properties of two Australian forests, one dry and open, the other moist and dense, were quantified and compared. Rates of reverberative decay and their dependence upon frequency were measured, together with the frequency-attenuating, or spectral, effects of sound propagation.

A. The field sites: Box-ironbark and mountain ash forests

Experimental trials were conducted in two distinctly different types of forest within southeastern Australia, the characteristics of which are compared in Table I. The first set of trials was carried out in relatively dry and open box-ironbark forest near Maldon, Victoria, approximately 125 km northwest of Melbourne. Three sites were selected in this forest (designated as box1, box2, and box3), chosen to have vegetation densities as uniform as possible. To examine repeatability, two identical trials (designated box1a and box1b) were conducted in one of the sites on different days. The second set of trials was within much moister and denser mountain ash forest, near Toolangi, Victoria, approximately 65 km northeast of Melbourne. Two of the trial sites here were on relatively flat ground. In the first of these the trials were repeated (ash1a and ash1b), while the remaining site (ash2) was in a slightly different location in the same forest stand. A third trial site (ash3) was in the bottom of a very shallow valley containing a small stream.

Tree densities and trunk diameters at breast height were measured within ten 10×10-m plots at each site; the results are summarized in Table II. The two forests are structurally rather different: the dominant mountain ash trees of the moist forest are much larger and far more widely spaced than those of the box-ironbark forest, with smooth-barked, columnar trunks unbroken for the first 20–25 m from the ground; the trunks of trees in the box-ironbark forest, on the other hand, are rough-barked and free of side branches for only the first 5–10 m. Although the relatively large trees of the mountain ash forest cover a lower proportion of the ground surface, that forest is structurally complex, with abundant shrubs and bushes occupying the lower vegetation strata; in contrast, the box-ironbark forest is structurally simple, with little vegetation other than the trees.

TABLE II. Quantitative description of the forests within which the experiments were conducted.

Forest	Tree densities[a] (per 100 m^2)	Trunk diameters[a,b] (cm)	% ground occupied by trees
Box-ironbark	13.90±11.09	13.5±6.0	5.9
Mountain ash	2.5±2.2	36.8±28.8	2.9

[a]Figures are mean±standard deviation.
[b]Measured at breast height.

The trials within the box-ironbark forest were conducted during August–September 1999. Maximum temperatures on all days were between 20° and 25 °C. Those within the mountain ash forest were conducted in April 2001, with maximum temperatures on all days of around 15 °C. (Such differences in temperature between the two forests are typical of most of spring and autumn.) To reduce the potential effects of scattering by turbulence and refraction by temperature stratification, all experiments were conducted between late morning and midafternoon on days of calm and stable weather.

B. Field techniques

Impulsive sounds were generated by bursting rubber balloons (''Meteor'' brand, made in Korea, 25-cm diameter) with a pin, after inflation to a fixed diameter of 22 cm. The sound-pressure level of bursts of these balloons was measured within an anechoic chamber, using a Bruel & Kjaer sound-level meter type 2209 and a 1/2-in. Bruel & Kjaer condenser microphone type 4165 placed 1.1 m from the balloons. The mean sound-pressure level was 133.2 ± 1.6 dB ($n=12$). Examination of digital recordings of these bursts, sampled at 44.1 kHz, always showed almost all of the decay (>80 dB) occurring within the first two sampling points, with levels varying <5 dB thereafter. The spectra of these bursts may therefore be presumed flat, at least below 20 kHz.

Investigations were made of the effects of three factors on rates of reverberative decay and spectral characteristics: source (balloon) height (h_s), receiver (microphone) height (h_r), and source–receiver distance (d). The effect of frequency (f) on reverberation rates was also investigated, as

described below. Four different source and receiver heights were used: 0.5, 1, 2, and 4 m. Within the box-ironbark forest, five distances were used (20, 40, 60, 80, and 100 m), resulting in $4\times4\times5=80$ balloon bursts for each experimental trial, 320 in total. Within the mountain ash forest attenuation proved to be much greater, resulting in amplitudes at 100 m being too low for analysis, so only four distances were used; $4\times4\times4=64$ balloons were burst for each trial, 256 in total. The overall total was 576 balloon bursts.

A Sennheiser ME20 omnidirectional microphone was set on a tripod at the required height. The balloons were tied to a rope slung over a tree branch, and burst with a pin attached to the end of a long, thin pole, so that the person bursting the balloons was always about 2 m distant. Balloon bursts were recorded with a Sony TCD-D8 DAT Walkman recorder at a sampling frequency of 44 100 Hz. Sensitivities of both the microphone and recorder varied <1 dB over the range of frequencies examined, and the total dynamic range exceeded 60 dB.

C. Data reduction

Signals were transferred to a Macintosh G3 computer for analyses, using a standard 3.5-mm audio input, and sampling at 22 050 Hz. The signals were analyzed with a program written using MATLAB v.5.2.1 (The Mathworks, Inc.).

1. Measurement of reverberative decay rates

The signals were filtered into ten contiguous frequency bands, giving 5760 estimates of reverberative decay rate. The filters were 500-point finite impulse response bandpass filters ($>$60-dB rejection), using Hamming windows; they were spaced at half-octave intervals (excepting the first band, which is a full octave) between frequencies of 200, 400, 566, 800, 1131, 1600, 2263, 3200, 4525, 6400, and 9051 Hz. For subsequent analyses, the representative frequencies for each band were taken at the respective quarter-octave points.

Amplitude envelopes (AEs) of each of the ten filtered signals were generated via a Hilbert transform filter. Because the transient response of the recording equipment is unknown, and may be unreliable in accurately recording such highly impulsive sounds as balloon bursts very near the peak of the burst, the initial peak of the AE was ignored, and the lower limit of the integration was set at the position of the peak immediately following this. The time difference was always very short, some tens of sample points, though the amplitude difference was often large. However, the difference between subsequent peaks proved to be much less, as illustrated in Fig. 1.

Rates of reverberative decay (in dB/s) were measured both directly, using a least-squares linear regression, and by using the integrated impulse response (IIR) technique (Schroeder, 1965). The problem of establishing an appropriate upper limit of integration for the IIR technique (Xiang, 1995) was overcome by using a two-piece linear regression: one piece describing the reverberative decay, the other describing the (presumably constant) background noise. The total residual error of this two-piece regression was examined as the joining point was moved throughout the signal. The upper limit was taken at the point of minimal error, plus

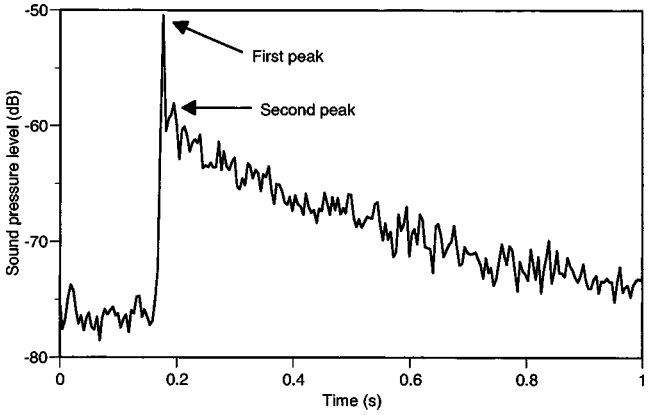


FIG. 1. Sample reverberative decay showing the difference in amplitude between the first and second peaks. All analyses begin at this second peak. The decay continues beyond the right-hand side of this figure, but has been truncated here.

a portion of background noise such that the slope of the reverse integral over this portion equaled the slope of the direct regression. For a perfectly exponential decay, the decay rate of the IIR decay equals that of the direct decay (Schroeder, 1965), so the degree of agreement between the two rates provides direct evidence for the existence of exponential decay. Furthermore, as an indication of possible systematic patterns of nonlinearity, a second-order polynomial regression was also performed upon the direct decay curves, and the curvature (in dB/s^2) extracted for subsequent analysis. (The slopes of these second-order curves were not examined.)

Even though very few signals began their decay at less than 40 dB above background, the measured signal levels as the signal decays to low signal-to-noise (SNR) ratios may be inaccurate, and this may affect regression results. To examine the degree of such effects, 20 signals were reanalyzed using a three-piece linear regression in place of the two-piece regression, with one piece describing the initial (linear) decay, one piece describing the (possibly inaccurate) signal decay during the low-SNR stage, and one piece describing the final, constant background noise level. Rates of decay measured from the first of these three pieces did not differ significantly from those obtained using two-piece regressions, supporting the accuracy of the latter.

2. Spectral analyses

Impulsive sources, having negligible durations, can be assumed to emit equal power at all frequencies, with the spectral properties changed during propagation by both differential frequency attenuation and the frequency-dependent nature of reverberation. In order to reduce this latter factor as much as possible, spectral analyses were centred close to the actual burst. A 512-point fast Fourier transform (duration $=23.1$ ms) was calculated over the portion of signal commencing at the time of the second peak (as described above). Although this does not preclude frequency-dependent effects of reverberation, inspection of recordings of balloon bursts taken 1 m from the source revealed generally flat spectra, indicating that such effects are greatly reduced. Furthermore, the only other method for measuring spectral effects, via

TABLE III. The performance of the MLRs for rates of reverberative decay within the two forests. All p values are <0.001.

Forest	Number of trials	Number of bursts pertrial	Range of R^2 values from MLRs	R^2 value after pooling all trials	Total number of bursts
Box-ironbark	4	~770	0.63–0.82	0.66	3083
Mountain ash	4	~620	0.57–0.77	0.50	2520

interrupted tone bursts, unavoidably includes the full frequency-dependent effects of reverberation.

For each experiment, the total spectral power at $d = 20$ m and $h_s = h_r = 4$ m was obtained, and this amount subtracted from all sound powers. All experiments were thereby internally calibrated to 0 dB at this reference configuration. To characterize the spectra, second-order regressions were performed upon spectral power versus frequency, and three primary parameters extracted. The first was the *spectral intercept at zero frequency* (in dB); this represents the maximal spectral power level, and provides a measure of overall attenuation. The others were the *power-frequency slope* (in dB/Hz), and the *curvature* (in dB/Hz^2). The values of spectral intercept were found to be strongly correlated with the values of total signal power ($R^2 = 0.72$, $n = 517$, $p < 0.001$), and independent of both the slope and curvature of spectra ($R^2 = 0.01$ and 0.06, respectively). Statistical analyses (see below) of the total signal power were also conducted, with results very similar to those obtained using the spectral intercept. However, the spectral intercept facilitates subsequent analyses, and is the measure of overall attenuation discussed below.

A single measure to characterize the *shapes* of a spectrum is obtained by calculating, for each value of slope and curvature, the maximal power range of the spectrum in dB: this is the difference between the maximal and minimal spectral powers, calculated from the quadratic polynomial regression. This maximal power range thus quantifies the relative extent of high-frequency attenuation, with greater values corresponding to relatively greater attenuation of higher frequencies.

D. Statistical analyses

Multiple linear regression (MLR) analyses were used because they provide a measure of the relative influence of each of the independent variables on the dependent variables. Second-order effects can be considered in MLRs by including the products of factors as separate terms. [Analyses of variance (ANOVAs) were also performed, with *all* results being highly significant ($p < 0.001$), but the relative influence of the independent variables is more directly quantified through MLRs.] Outlying points were expected in the data, caused by such factors as extraneous bird vocalizations during the recordings. For each trial, the distributions of residuals generated from an MLR were examined, and any cases further than two standard deviations from the center of the distribution excluded from further analyses. From the 5760 measurements of decay rate, a total of 157 decays (2.7%) was removed in this way. Doing so considerably improved subsequent regression fits. Out of the 576 spectral estimates, 59—or 10% of the total—were identified as outliers. This rather high proportion occurred because bird vocalizations, which occupy a narrow spectral range, were relatively frequent and greatly distort the overall shapes of the spectra; removal of this 10% vastly improved subsequent regressions.

It is also recognized that the dependent variables may not necessarily be linear functions of the independent variables, so each independent variable is included twice: in linearly- and logarithmically scaled forms. The MLR technique then allows for possible linear dependence on these logarithmically scaled variables; the two forms of each variable will tend to be selected according to the extent to which they *independently* explain variations in the dependent variables. (Note that this procedure may reduce the coefficients of determination.)

Eight first-order variables were included in each regression analysis, being the linearly- and logarithmically scaled versions of d, h_s, h_r, and f. Second-order variables were obtained by multiplying either the linearly- or logarithmically scaled version of each variable by both versions of all other variables (e.g., $d \times h_s$ and $d \times \log(h_s)$, but not $d \times \log(d)$). There was a total of 24 second-order variables in each analysis, and 32 variables altogether. A stepwise selection procedure was employed ($p_{in} = 0.05$, $p_{out} = 0.10$) to select those variables having significant relationships with the independent variables. Other selection procedures were examined, with results found to be essentially the same as for the present procedure. Second-order variables in most cases made a considerable improvement to the model, but were included even when they did not to ensure consistency between analyses.

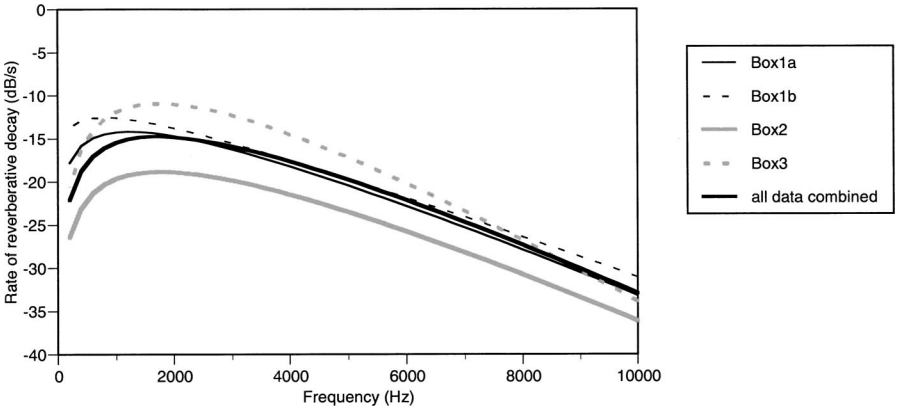


FIG. 2. Rates of reverberative decay as a function of frequency from the four experimental trials within the box-ironbark forest, and rates obtained with data from all trials combined. There is a strong similarity between the repeated trials (box1a and box1b). Variation within trials was always much less than variation between trials (all ANOVAs gave $p < 0.001$, $n \sim 640$), and is not shown here.

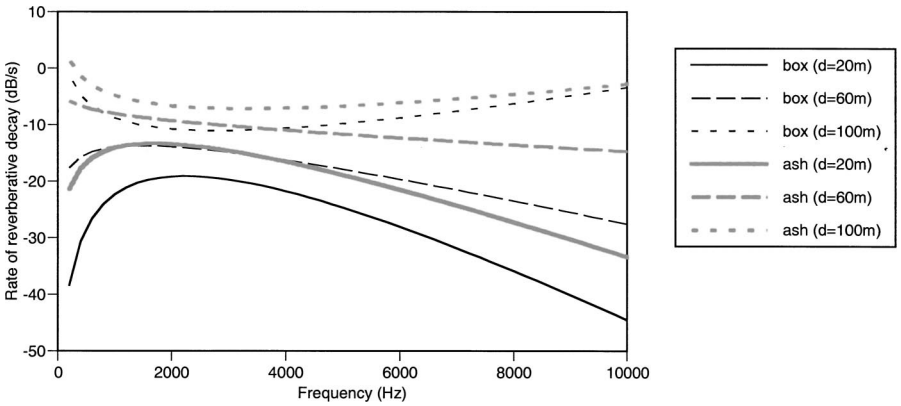


FIG. 3. Rates of reverberative decay within the two types of forest, as a function of frequency for three source–receiver distances. Lines for similar distances in the two forests are close to parallel, exhibiting the same qualitative aspects.

III. RESULTS

A. Reverberative decay rates

Interestingly, rates of reverberative decay were almost identical whether measured with the IIR method or from direct regressions on decay curves. In fact, the lowest correlation coefficient between the two sets of measurements for a single trial was 0.978, while almost all others were above 0.99 ($n \geqslant 640$ in all cases, $p < 0.001$). Therefore, in the following results, all references to, and illustrations of, rates of decay refer to directly measured rates; the results for IIR decay rates can be taken to be almost identical.

MLR analyses were first calculated separately for each experimental trial. The rates of decay from the four trials within each forest were then pooled, and a single MLR performed to generate an overall model of reverberative decay for each forest type. Measures of fit were provided by the adjusted R^2 values from the multiple-linear regressions, as shown in Table III. In the following results, unless otherwise specified, default values for the configurational variables are used. These are $h_s = h_r = 2$ m, $d = 50$ m, and $f = 5$ kHz. Note also that, unless explicitly mentioned as nonsignificant, all MLR results can be assumed to have been significant at $p < 0.001$.

Figure 2 shows the rate of reverberative decay as a function of frequency from the four trials within the box-ironbark forest, and for the pooled data. In both cases, results from the two identical repeats (box1a and box1b) are more similar to one another than to results from any other locations, suggesting a high degree of repeatability of these measurements. Rates of reverberative decay from the four trials within the structurally denser mountain ash forest were similarly distributed, and are not shown here. The MLR results obtained from pooling each group of four trials were then used to generate all subsequent results. Figure 3 shows the effects of frequency on rates of reverberative decay for three distances ($d = 20$, 60, and 100 m) within both types of forest. There is a striking similarity between these two types of forest in the dependence of reverberative decay rates on frequency. There is a central region of frequencies between 1 and 3 kHz, wherein reverberation varies much less with distance than for frequencies outside this region. This is the frequency region of *slowest* reverberative decay at close distances (i.e., 20 m), and the region of *fastest* decay at far distances (i.e., 100 m). Note also the considerably slower rates of decay within the mountain ash forest.

Changes in the heights of the source and receiver have a much smaller effect on the rates of reverberative decay than changes in either distance or frequency and, within mountain ash forest, changes in receiver height do not have any significant effect. The significant effects for the box-ironbark forest are shown in Fig. 4. Where a source is close to the ground, changes in the height of a receiver have a relatively large effect, whereas if a source is higher above the ground, changes in the height of a receiver have a lesser effect. Furthermore, changes in the receiver height further depend upon distance. Increases in the height of a receiver produce slower reverberative decay at shorter distances, but produce faster reverberative decay at the farthest distance of 100 m.

The final parameter extracted from each reverberative decay is the *curvature* of the decay. MLR results for these are shown in Fig. 5. From these it can be seen that curvature is always positive, suggesting an initially rapid decay fol-

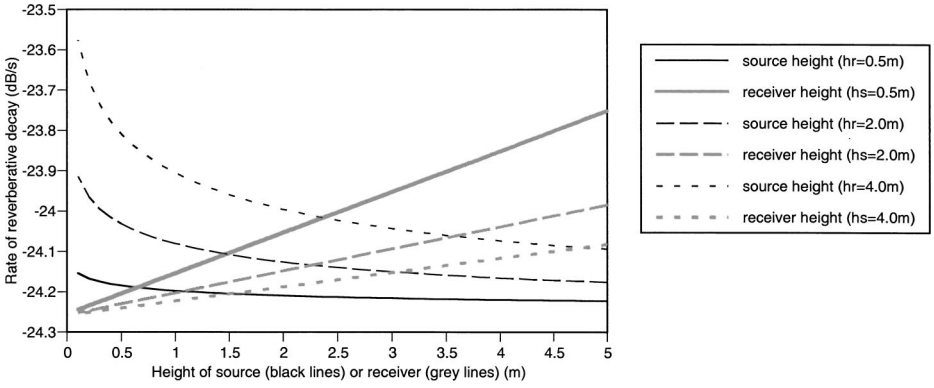


FIG. 4. Rates of reverberative decay as a function of the heights of both source and receiver when they are separated by a distance of 20 m with box-ironbark forest. Different source heights were examined while holding the receiver at a height of 2 m, and vice versa. Increasing source height leads to faster reverberative decay, while increasing receiver height slows the decay of reverberation. Changes in both cases are very slight, however.

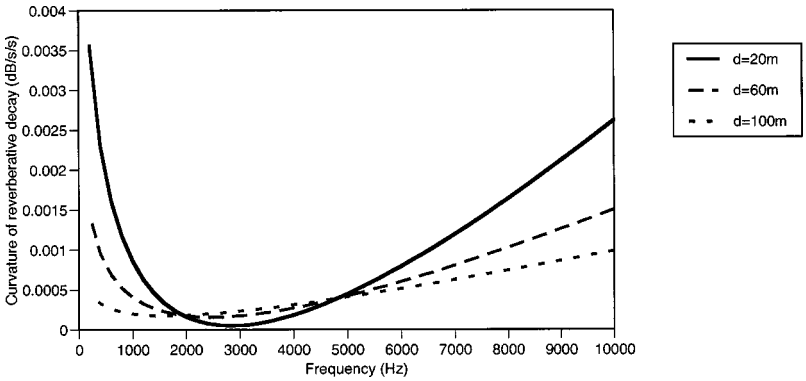


FIG. 5. Curvature of reverberative decay as a function of frequency within the box-ironbark forest. Note the central band of frequencies (2–5 kHz) within which reverberative decay is more linear than for other frequencies. This band corresponds to the band evident in Fig. 3, within which rates of reverberative decay vary relatively little with distance. Results from the mountain ash forest were very similar.

lowed by a progressively less rapid decay, although the values for curvature are very small, indicative of highly linear decay (in a logarithmic scale). Furthermore, the patterns evident in this figure are what might be expected if the *curvature* were negatively correlated with the (linear) *rate* of reverberative decay, as can be seen by comparing Fig. 5 with Fig. 3. Indeed, the correlation coefficient between all values of the slope and curvature of reverberative decay is -0.52 ($n=5603$), which may explain why values of curvature from the box-ironbark forest are generally greater than those from the mountain ash forest. These results suggest the existence of a band of frequencies, centered around 2–3 kHz, within which reverberative decay is essentially exponential (i.e., linear in a dB scale), while frequencies further away from this band exhibit progressively less exponential and more upwardly curved decay. This effect is much less pronounced at greater distances, especially within the more structurally complex mountain ash forest. These positive values of curvature, and their strong relationship with the slopes, do not seem to be affected by the decay reaching the dynamic limit of the recording equipment: the full dynamic range of the recorded decays (measured in dB) is not related to either the slope or the curvature of the decays (R^2 values <0.005 in both cases; $n=5603$).

B. Spectral analyses and attenuation

MLRs on the spectral parameters were initially performed on the individual trials to permit inspection of the amount of variation between trials, and then the data were pooled before performing final regressions to characterize spectra from the two types of forest. Variation both within and between trials proved to be similar to the respective variations in rates of reverberative decay described above; therefore, only the results from the pooled data sets are included here. The performance of MLRs for spectral parameters is summarized in Table IV.

1. Spectral intercepts and overall attenuation

MLRs for the spectral intercepts are shown as a function of source–receiver distance in Fig. 6. A strong increase in attenuation with increasing distance is immediately apparent such that, in both types of forest, attenuation increases by about 35 dB over the range of distances examined (20–100 m), with attenuation always greater in the mountain ash forest than in the box-ironbark forest. Spectral intercepts were also significantly related to both source and receiver heights in both types of forest, as shown in Fig. 7.

2. Spectral shapes and differential frequency attenuation

Distance is by far the most influential variable in explaining patterns of variation in both spectral slope and curvature for both types of forest, and the resultant measure of maximal power range. Figure 8 shows the relationships between distance and maximal spectral range for both types of forest. Values for the mountain ash forest are generally greater than those for the box-ironbark forest. However, distance has less effect on the degree of high-frequency attenuation in the mountain ash forest than it does in the box-ironbark forest, although increasing distance in both types of forest generally leads to lower values of maximal spectral range. This implies that high-frequency attenuation becomes relatively less pronounced with increasing distance, although it should be kept in mind that the spectral intercept decreases markedly with increasing distance. This decrease in overall sound-pressure level with distance, evident in Fig. 6, may also lead to a lower dynamic range being recorded for each balloon burst, which may possibly contribute to the general pattern of decrease in maximal spectral power range with distance observed in Fig. 8. However, the greater maximal spectral power range in the mountain ash forest, which also has greater overall attenuation, is inconsistent with this explanation, suggesting that the results for maximal spectral range are not confounded in this manner.

Spectral slopes and curvatures within the mountain ash forest show no significant dependence on either source or receiver height, while those within the box-ironbark forest

TABLE IV. The performance of the MLRs for rates of parameters of spectral shape within the two forests. All p-values are <0.001 unless otherwise noted.

Forest	Number of spectra	R^2 values for spectral intercept	R^2 values for spectral slope	R^2 values for spectral curvature[a]
Box-ironbark	320	0.85	0.29	0.29
Mountain ash	197	0.84	0.20[b]	0.26

[a]Spectral curvature was closely related to spectral slope ($R=-0.87$, $n=517$, $p<0.001$).
[b]$p=0.032$.

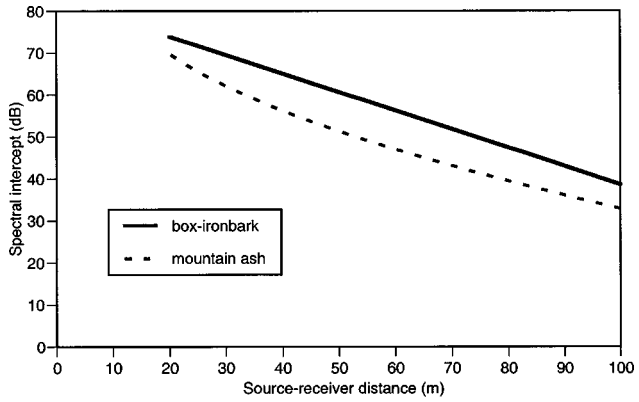


FIG. 6. Spectral intercept as a function of source–receiver distance in the two types of forest. Attenuation was always greater in the mountain ash forest, but the rates of change with distance were similar in the two forests. Note that the dB values are arbitrarily scaled.

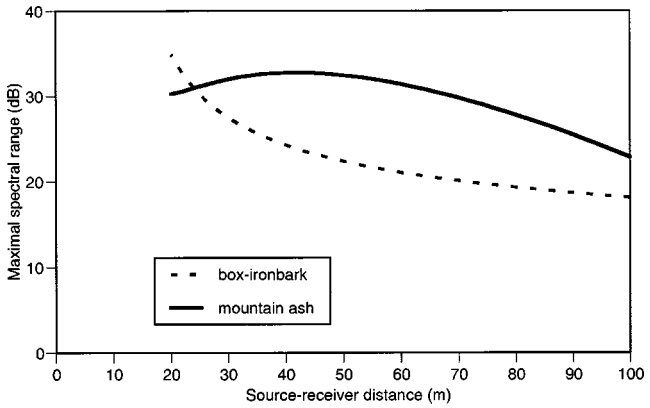


FIG. 8. Maximal spectral range as a function of distance between source and receiver in the two forests. Note that spectral ranges are generally greater in the mountain ash forest, indicating relatively greater differences between attenuation at low and high frequencies and thus relatively greater high-frequency attenuation.

show a significant dependence that is very similar to the dependence of spectral intercepts, shown in Fig. 7.

IV. DISCUSSION

A. Measuring rates of reverberative decay within forests

The set of experiments described here shows that reverberative decay can be appreciable within forested environments, and that the measured decay rates show a significant dependence on distance, the heights of the source and receiver, and frequency. Very interestingly, the rates of reverberative decay measured directly through linear regressions on the decay curves correlate almost perfectly with the decay rates measured using the IIR technique (Schroeder, 1965). This very strong agreement seems to provide strong evidence that reverberation decays exponentially within the forested environments examined.

B. Frequency dependence of reverberative decay

One of the most interesting findings from this series of experiments is the dependence of the rate of reverberative decay on frequency, and the way in which that dependence changes with the distance between the source and receiver. In both types of forest at close distances, there exists a central region of frequencies between 1 and 3 kHz in which reverberation decays markedly more slowly than it does for frequencies outside this region. As distance increases, the differences in reverberative decay rate with frequency that distinguish this region are lessened until they are eventually reverse at a distance of 100 m, so that this central region exhibits somewhat faster rates of reverberative decay than do frequencies outside the region. Aside from the expected tendency for generally slower reverberative decay in the structurally denser (mountain ash) forest, the dependence of reverberative decay rate on frequency, and the way it changes with distance, is almost identical within the two types of forest.

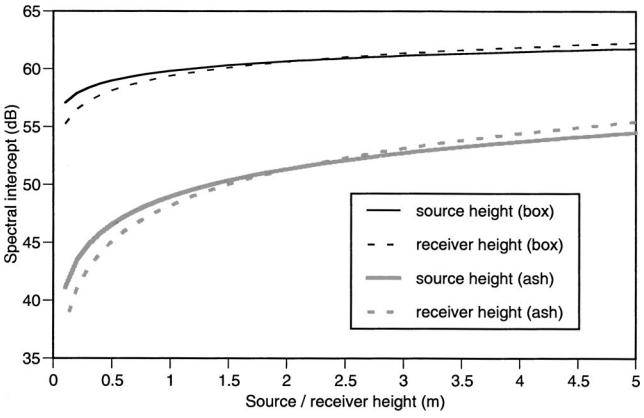


FIG. 7. Spectral intercept as a function of the heights of both source and receiver. Increasing height at either end reduces attenuation, with the effect relatively more pronounced for the receiver height than the source height in both forests. Furthermore, changes in height have a relatively greater effect on attenuation in the mountain ash forest than in the box-ironbark forest.

The frequencies that delimit this central region correspond to wavelengths between approximately 3 and 9 cm. This range encompasses the dimensions of most individual leaves within the two types of forest, as well as the typical diameters of most tree branches. Examination of Fig. 3 reveals that, as distance changes, the rates of reverberative decay change more for frequencies outside this region than they do for frequencies within it. This may be because foliage effectively produces a more diffuse acoustic field within this region, reducing the degree of dependence of reverberative decay rate on distance that might otherwise occur. The dependence of the curvature of reverberative decay on frequency, illustrated in Fig. 5, supports this by showing that reverberation is most nearly exponential around this central band of frequencies, curving ever more upward as frequencies move further away from around 3 kHz.

This finding is very interesting in the light of discussions of the importance of *ranging* (i.e., estimation of distance based on acoustical degradation) in animal communication, especially for birds (Holland *et al.*, 1998; Mathevon, 1998; McGregor and Krebs, 1984; Morton, 1986). The range of frequencies between 1 and 3 kHz corresponds approximately to those used by many forest bird species (although there are also many species which vocalize at frequencies above 3

kHz). If the kind of acoustical degradation imposed through reverberation is to be used for ranging, birds are likely to find it more effective to use the frequencies outside this region that exhibit a strong relationship between distance and rates of reverberative decay. Conversely, if ability to range a song were disadvantageous, this central range of frequencies may offer distinct advantages. Examination of relationships between rates of reverberation and the frequencies used in birdsong may be a fruitful way of examining such issues in future work.

C. Source and receiver heights

The dependence of the rates of reverberative decay on source and receiver heights shown in Fig. 4 may shed some light onto the debate over the importance of song perches for singing birds (Holland *et al.*, 1998; Hunter, 1980; Lemon *et al.*, 1981). The fact that source and receiver heights are generally more influential in the box-ironbark forest is not surprising given that the height range used spans proportionally more of the total vegetation heights within that forest. In both forests, broadcasting signals from height always leads to reverberation decaying faster, whereas the effects of listening at height, at least within the box-ironbark forest, also further depend upon distance. Increases in the height of a listener only produce faster reverberative decay at the farthest distance examined (100 m). However, even at these distances, increasing source height has a relatively greater effect on the rate of reverberative decay than does increasing receiver height.

The dependence of overall attenuation on distance is precisely as expected, with attenuation increasing with distance in both forest types, and tending to be greater in the mountain ash forest. However, attenuation is reduced more by increases in receiver height than in source height, with effects similar in both forests. This contrasts with the effects of source and receiver heights on the rate of reverberative decay. Increasing the height of the source seems to be consistently more effective at reducing the impact of reverberation than any change in receiver height.

This has interesting implications for the detectability and clarity of acoustic signals emitted within these forests. Signal *detectability* will be affected to a far greater extent by attenuation than reverberation. However, reverberation may alter the structure or clarity of a signal so much that, even if it is detected, it may not be recognized as informative. The detectability of signals may be better improved through increases in the height of receivers than in the height of senders. In contrast, the *clarity* of a signal may be better improved through increases in sender height than in receiver height. Given that both detectability and clarity are needed for a sound signal to be effective in animal communication, detection is the primary function: a signal must be detected first before it can be recognized, but not vice versa. This suggests that perching at height among birds is as important in listening to other birds' songs as it is to effective singing.

The debate over the relative benefits to senders and receivers of increasing height has reached little accord (Dabelsteen *et al.*, 1993; Hunter, 1980; Lemon *et al.*, 1981; Mathevon *et al.*, 1996), and these results perhaps offer some explanation for the confusion. The relative merits may depend on whether one considers detectability or clarity of a signal more important in a given situation. The results of Dabelsteen *et al.* (1993) suggest that receiver height is far more important than source height in determining the degradation of songs. Their results were elaborated in the more detailed study of Holland *et al.* (1998), who found that increases in receiver height were far more important than increases in source height in reducing signal-to-noise ratio, and that increases in height at either end were close to equally important in decreasing excess attenuation. The present results are entirely consistent with these conclusions, but provide unambiguous pictures of the separate effects upon reverberation and attenuation of raising the heights of sources and receivers.

It should also be remembered that the ranges of heights examined within these experiments were the same in both types of forest. Within the box-ironbark forests, these heights span a range from approximately ground level up to the beginning of the upper layer of foliage, which probably encompasses much of the height range likely to be used for communication between birds within these forests. However, within the mountain ash forests, the maximal height of 4 m is barely above the shrub layer, and far below the lowest levels of foliage of the dominant mountain ash trees. These have almost no foliage below at least 20 m, and the range of heights used by many birds communicating within these forests is greater than the range examined here. It would perhaps be more reasonable to rescale the heights used within the box-ironbark forest (0.5 to 4 m), where trees are around 12 m high, to a scale equivalent to a tree height of 45 m (2 to 15 m). Conducting these experiments at a height of 15 m, however, would be far more technically difficult, and was beyond the scope of the present work.

V. CONCLUSIONS

These results provide convincing evidence of exponential reverberative decay within the two forests studied here, particularly within a broad band of frequencies between 1 and 3 kHz. One striking finding is that rates of decay change much less with distance within this band than for any other frequencies, and that this pattern is very similar in both forests. These frequencies correspond to wavelengths of approximately the same dimensions as the foliage of the tree species within the forests, suggesting that similar conditions may occur within many other forests elsewhere.

However the acoustic environments of the two forests exhibit other striking differences that may have important implications for acoustic communication. Both attenuation and reverberation depend very strongly on the distance between the source and receiver, and to a lesser extent upon the heights of the source and receiver. Increasing the source height is generally more effective at reducing reverberation than increasing the receiver height, in contrast to attenuation which is better reduced by increasing the receiver height than the source height. That these patterns hold in both forests suggests that this may be a general feature of acoustic communication in forested environments.

These experiments offer a relatively simple way to examine several different aspects of acoustic propagation within many complex, natural environments. Importantly, they provide a method by which the acoustic environments of any forests may be quantified and directly compared, and illustrate the potential value of such comparisons.

ACKNOWLEDGMENTS

I am grateful to David Morgan for extensive advice on many aspects of this work, including advice on this manuscript. Murray Littlejohn and David Paul generously loaned much of the equipment used for the field work.

Aylor, D. (**1972**). ''Sound transmission through vegetation in relation to leaf area density, leaf width, and breadth of canopy,'' J. Acoust. Soc. Am. **51**, 411–414.

Badyaev, A. V., and Leaf, E. S. (**1997**). ''Habitat associations of song characteristics in *Phylloscopus* and *Hippolais* warblers,'' Auk **114**, 40–46.

Chappius, C. (**1971**). ''Un example de l'influence du milieu sur les emissions vocales des oiseax: L'evolution des chants en foret equatoriale,'' (An example of the influence of the environment of the vocalizations of birds: The evolution of songs in equatorial forests.) Terre Vie **25**, 183–202.

Dabelsteen, T., Larson, O. N., and Pedersen, S. B. (**1993**). ''Habitat-induced degradation of sound signals: Quantifying the effects on communication sounds and bird location of blur ratio, excess attenuation, and signal-to-noise ratio in blackbird song,'' J. Acoust. Soc. Am. **93**, 2206–2220.

Holland, J., Dabelsteen, T., and Pedersen, S. B. (**1998**). ''Degradation of wren *troglodytes troglodytes* song—implications for information transfer and ranging,'' J. Acoust. Soc. Am. **103**, 2154–2166.

Huisman, W. T., and Attenborough, K. (**1991**). ''Reverberation and attenuation in a pine forest,'' J. Acoust. Soc. Am. **90**, 2664–2677.

Hunter, M. L. (**1980**). ''Microhabitat selection for singing and other behaviour in great tits, *Parus major*: Some visual and acoustical considerations,'' Anim. Behav. **28**, 468–475.

Kuttruff, H. (**1967**). ''Uber Nachhall in Medien mit unregelmassig verteilten Streuzentren, insbesondere in Hallraumen mit aufgehangten Streuelementen,'' (On reverberation in media with irregularly distributed scattering centers, in particular rooms with suspended scattering elements) Acustica **18**, 131–143.

Kuttruff, H. (**2000**). *Room Acoustics*, 4th ed. (Spon, London).

Lemon, R. E., Struger, J., and Lechowicz, M. J. (**1981**). ''Song features and singing heights of American warblers: Maximization or optimization of distance?'' J. Acoust. Soc. Am. **69**, 1169–1176.

Martens, M. J. M. (**1980**). ''Foliage as a low pass filter: Experiments with model forests in an anechoic chamber,'' J. Acoust. Soc. Am. **67**, 66–72.

Martens, M. J. M., and Michelsen, A. (**1981**). ''Absorption of acoustic energy by plant leaves,'' J. Acoust. Soc. Am. **69**, 303–306.

Mathevon, N. (**1998**). ''Degraded temporal sound features as a function of distance and potential as cues for ranging in birds,'' Bioacoustics **9**, 17–33.

Mathevon, N., Aubin, T., and Dabelsteen, T. (**1996**). ''Song degradation during propagation—Importance of song post for the wren *Troglodytes troglodytes*,'' Ethnology **102**, 397–412.

McGregor, P. K., and Krebs, J. R. (**1984**). ''Sound degradation as a distance cue in great tit (*Parus major*) song,'' Behav. Ecol. Sociobiol. **16**, 49–56.

Morton, E. S. (**1975**). ''Ecological sources of selection on avian sounds,'' Am. Nat. **109**, 17–34.

Morton, E. S. (**1986**). ''Predictions from the ranging hypothesis for the evolution of long-distance signals in birds,'' Behaviour **99**, 65–86.

Naguib, 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. (**2003**). ''Reverberation of rapid and slow trills: Implications for signal adaptations to long-range communication,'' J. Acoust. Soc. Am. **113**, 1749–1756.

Polack, J. D. (**1992**). ''Modifying chambers to play billiards: The foundations of reverberation theory,'' Acustica **76**, 257–272.

Price, M. A., Attenborough, K., and Heap, N. W. (**1988**). ''Sound attenuation through trees: Measurements and models,'' J. Acoust. Soc. Am. **84**, 1836–1844.

Richards, D. G., and Wiley, R. H. (**1980**). ''Reverberations and amplitude fluctuations in the propagation of sound in a forest: Implications for animal communication,'' Am. Nat. **115**, 381–399.

Schroeder, M. R. (**1965**). ''New method of measuring reverberation time,'' J. Acoust. Soc. Am. **37**, 409–412.

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 Godard, R. (**1996**). ''Ranging of conspecific songs by Kentucky warblers and its implications for interactions of territorial males,'' Behaviour **133**, 81–102.

Wiley, R. H., and Richards, D. G. (**1982**). ''Adaptations for acoustic communication in birds: Sound transmission and signal detection,'' in *Acoustic Communication in Birds 1*, edited by D. E. Kroodsma and E. H. Miller (Academic, New York), pp. 132–182.

Xiang, N. (**1995**). ''Evaluation of reverberation times using a nonlinear regression approach,'' J. Acoust. Soc. Am. **98**, 2112–2121.

Sex differences in the acoustic structure of vowel-like grunt vocalizations in baboons and their perceptual discrimination by baboon listeners

Drew Rendall[a)]
Department of Psychology and Neuroscience, University of Lethbridge, Lethbridge, AB T1K 3M4 Canada

Michael J. Owren
Department of Psychology, Cornell University, Ithaca, New York

Elise Weerts and Robert D. Hienz
Division of Behavioral Biology, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland

(Received 6 February 2003; revised 24 September 2003; accepted 3 November 2003)

This study quantifies sex differences in the acoustic structure of vowel-like grunt vocalizations in baboons (*Papio spp.*) and tests the basic perceptual discriminability of these differences to baboon listeners. Acoustic analyses were performed on 1028 grunts recorded from 27 adult baboons (11 males and 16 females) in southern Africa, focusing specifically on the fundamental frequency (F_0) and formant frequencies. The mean F_0 and the mean frequencies of the first three formants were all significantly lower in males than they were in females, more dramatically so for F_0. Experiments using standard psychophysical procedures subsequently tested the discriminability of adult male and adult female grunts. After learning to discriminate the grunt of one male from that of one female, five baboon subjects subsequently generalized this discrimination both to new call tokens from the same individuals and to grunts from novel males and females. These results are discussed in the context of both the possible vocal anatomical basis for sex differences in call structure and the potential perceptual mechanisms involved in their processing by listeners, particularly as these relate to analogous issues in human speech production and perception. [DOI: 10.1121/1.1635838]

PACS numbers: 43.80.Ka, 43.80.Lb, 43.70.Gr [WA] Pages: 411–421

I. INTRODUCTION

In this paper, we examine acoustic–structural differences in the grunt vocalizations produced by male and female baboons and test the perceptual discriminability of such differences to the animals. Sex differences in vocal communication are not uncommon in the animal world (Hauser, 1996). In some cases, they involve the production of distinctive calls, or patterns of calling, by one sex and not the other that seem to reflect consistent sex differences in the behavioral, motivational, physiological, or neurological systems subserving communication. One familiar example of this type of difference is the singing of many species of songbird, where males but not females produce distinctive vocalizations during territorial rivalry and female courtship. Here, the production of distinctive vocalizations by males appears to be traceable, at least in some species, to hormonal differences between males and females that affect the basic motivation to sing in addition to influencing the size and structure of specific nuclei in the brain that regulate the details of song production (reviewed in Hauser, 1996).

In other cases, males and females produce the same kinds of vocalizations in the same behavioral contexts, but the detailed acoustic structure of the calls they produce is noticeably different, often because of at least subtle differences between the sexes in the anatomical features involved in sound production. A prominent example of this type of difference is human speech. Both sexes produce the same speech sounds (the ones that are characteristic of whatever language they speak), but their detailed acoustic structure can nevertheless differ appreciably because of differences between the sexes in the size and shape of two independent but complementary features of vocal production anatomy, namely the vocal folds of the larynx and the airways of the vocal tract above the larynx (reviewed in Lieberman, 1986; Fitch, 2000). Acoustic–structural differences in the speech of adult males and females are most obvious in vowel sounds which clearly manifest both the vibratory action of the vocal folds and the resonance properties of the supralaryngeal vocal-tract cavities. A consistent sex difference in the length of the vocal folds results in a reliable difference in the rate at which the vocal folds naturally vibrate (the fundamental frequency of phonation, or F_0) during relaxed speech (Titze, 1989). This difference translates perceptually as a difference in the average pitch of vowel sounds, which is dramatically lower in adult males than in adult females. Other sex differences in the length and shape of the vocal-tract cavities above the larynx affect the resonance, or formant, structure of the vowel sounds produced (e.g., Fant, 1966), yielding frequencies that are again appreciably (though less dramatically) lower in adult males than they are in adult females

[a)]Electronic mail: d.rendall@uleth.ca

(e.g., Peterson and Barney, 1952; Hillenbrand *et al.*, 1995; Bachorowski and Owren, 1999).

These vocal anatomical differences between the sexes can themselves be traced to some extent to an overall difference in body size between males and females, a condition referred to as sexual dimorphism (reviewed in Fitch and Giedd, 1999). Human males are, on average, taller and heavier than females, and, as a result, many features of bodily anatomy, including vocal anatomy, are larger in males than they are in females. Although the precise nature of the relationships linking overall body size dimorphism, vocal anatomical dimorphism, and dimorphic features of the voice in humans is not entirely clear (e.g., Lass and Brown, 1978; Kunzel, 1989; van Dommelen, 1993; Fitch and Giedd, 1999; Collins, 2000), the acoustic–structural differences in the voices of males and females are obvious and represent an important source of speech variation. In fact, because many of the different vowels and other phonemes of language are determined in large part by their different patterns of vocal-tract resonance, the baseline differences in resonance structure between adult males and females are a critical source of speech variation that listener perceptual systems have to normalize in order to recover the linguistic content of utterances. Such normalization is hypothesized to involve a process that effectively compensates for the baseline sex differences in resonance structure in order to isolate that component which reflects phoneme identity (Ladefoged and Broadbent, 1957; Nearey, 1989). However, here again, the details of the process are not completely understood. Furthermore, it is not entirely clear whether such normalization represents a species-specific perceptual mechanism in humans specialized for processing speech sounds, or a more general capacity of the mammalian auditory system that functions widely to normalize potential sex differences in vocalizations (reviewed in Johnson and Mullenix, 1997; Trout, 2001).

We have recently begun to examine some of these issues through comparative research on baboons. As primates relatively closely related to humans, baboons provide a logical opportunity to trace potential continuity in the processes of human vocal production and perception, and perhaps also to help clarify some of the basic anatomical and acoustic–structural relationships involved. In fact, research on baboon communication has already revealed a variety of parallels to speech phenomena. Not only is general vocal-tract anatomy and function fundamentally similar in nonhuman primates and humans (Fitch and Hauser, 1995; Owren and Linker, 1995), but some calls produced by baboons in particular are remarkably speech-like. For example, the most common vocalization produced by baboons—a harmonically rich grunt—is structurally analogous to human vowel sounds. Several different authors have described a clear, stable F_0 and a series of prominent spectral peaks, each encompassing several harmonics that appear to reflect the independent action of stable vocal-fold vibration and subsequent spectral shaping by the resonance characteristics of the supralaryngeal airways (Zhinkin, 1963; Andrew, 1976; Richman, 1976). Furthermore, recent detailed analyses by Owren *et al.* (1997) and Rendall (2003) have found that the mean F_0 of female grunts is approximately 115 Hz, and the means of the first three formant frequencies are approximately 450, 1440, and 2700 Hz, respectively. These values match closely mean F_0 (117 Hz) and formant values (502, 1420, 2386 Hz) for vowels produced by English-speaking males (Baken, 1987). This similarity is illustrated in Fig. 1, which plots the first two formant frequencies of a sample of female baboon grunts within the standard vowel space represented by the first two formant frequencies of American English vowels. The grunts occupy a central position in this plot, overlapping most with the space occupied by the ''schwa'' vowel that is often regarded as a ''neutral'' sound that involves little or no active vocal-tract articulation.

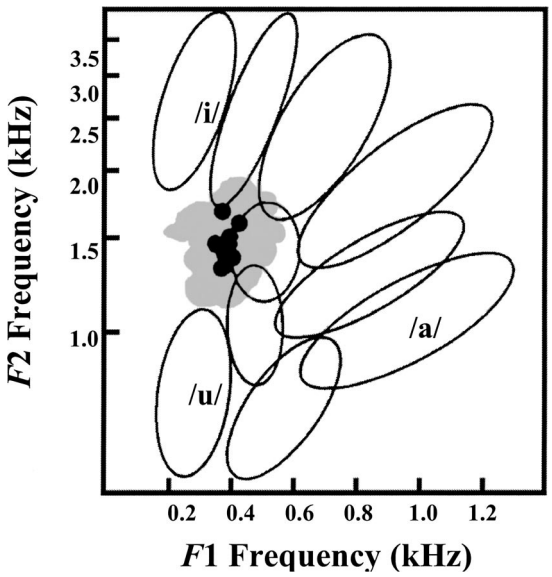


FIG. 1. Mean frequency values of $F1$ and $F2$ for each of nine adult female baboons (black circles) plotted within the American-English $F1/F2$ vowel-space mapping of Peterson and Barney (1952). The larger shaded (gray) area represents the entire range of variation in the grunt sample. The large, open ellipses represent the space occupied by individual vowel sounds. The point vowels /i/, /a/, and /u/ are labeled. (Redrawn after Owren *et al.*, 1997.)

This acoustic–structural convergence between baboon grunts and human vowels has been attributed to a likely similarity between the species in vocal-fold and vocal-tract lengths, the latter probably resulting from baboons' relatively long muzzle. By comparison to humans, baboons, like many nonhuman primates, have a short pharyngeal cavity resulting from a relatively high positioning of the larynx in the vocal tract (Negus, 1949; Lieberman, 1968; Fitch 2000). However, their developed snout contributes to a much longer oral cavity which appears to yield an effective overall vocal-tract length in adult female baboons that is very similar to that of human males (Owren *et al.*, 1997).

In addition, research suggests that baboons may produce subtly distinct variants of the grunt vocalization in different behavioral contexts, and, again like human vowels, the acoustic features that best distinguish these grunt variants pertain to vocal-tract resonance properties (Owren *et al.*, 1997). Specifically, grunts produced by females about to embark on a move across open terrain (so-called ''move'' grunts) can have a flatter spectral slope and a higher second formant frequency ($F2$F) compared to grunts produced in the context of social interactions involving the inspection

and handling of the young infants of other females in the group (so-called "infant" grunts), which have a steeper spectral slope and lower $F2F$. Natural sensitivity to these acoustic differences has been confirmed in experimental studies of the baboons conducted in the wild (Rendall *et al.*, 1999). General sensitivity to spectral resonance contrasts has also been confirmed in systematic laboratory perceptual experiments on baboons (Hienz and Brady, 1988). This work, and similar studies of closely related primate species (e.g., Sinnott and Kreiter, 1991; Sommers *et al.*, 1992), has shown frequency discrimination abilities for formant-type contrasts that are very similar to those of humans (Kewley-Port and Watson, 1994; Hienz, unpublished data).

Like female baboons, males also grunt regularly in the behavioral contexts just described. Although male grunts have not been formally described, preliminary acoustic analyses indicate that they are structurally the same as female grunts except that their F_0 and resonances appear to be appreciably lower. These acoustic–structural differences between them probably stem, as they do in humans, from differences in vocal-fold and vocal-tract lengths arising from sex differences in overall body size (Fitch and Giedd, 1999; Fischer *et al.*, 2002). Baboons are strongly sexually dimorphic, with adult males being nearly twice the size of adult females (Barrett and Henzi, 1997).

Taken together, these various findings recommend baboons for comparative research examining both the mechanistic basis of sex differences in call structure and the potential processes mediating the perception of such differences. The necessary first two steps in this work are to systematically quantify acoustic–structural differences in the grunts of adult males and females and then to test the basic discriminability of these differences to baboon listeners. This paper reports the results of these two research steps.

II. SEX DIFFERENCES IN THE ACOUSTIC STRUCTURE OF GRUNTS

A. Methods

1. Field site, study subjects, and vocalization recording

Vocalizations used in this study were collected during field research on chacma baboons (*Papio hamadrayas ursinus*, formerly *P. cynocephalus ursinus*) at two sites in southern Africa. One site was located in the Moremi Game Reserve in northern Botswana, which lies within the Okavango Delta. The other site was in the De Hoop Nature Reserve in the Western Cape Province of South Africa. Baboons at the two sites belong to the same species and subspecies that is distributed continuously throughout southern Africa (i.e., Namibia, Botswana, Zimbabwe, and South Africa). Subjects were the sexually and physically mature adult males (>8 years) and females (>7 years) of the main study group at each site. Both groups had been studied continuously for several years. As a result, the animals were fully habituated to human observers and high-quality recordings of their vocalizations could therefore be made at very close range (0.5–2.0 m). Vocalization recordings were made during a single protracted study period in Botswana (February 1996–March 1997) and during multiple shorter study periods in South Africa (December 2000; May–August 2001; January–May 2002).

TABLE I. Summary of the grunt samples from individual male ($N=11$) and female ($N=16$) baboons. Each baboon is identified with a letter/number code, and the number of grunts used for each individual is indicated in brackets.

Males		Females		
AP (50)	PI (12)	AI (21)	BT (89)	SH (94)
AU (28)	RF (20)	AL1 (62)	CT (69)	SS (89)
CS (46)	RO (6)	AL2 (23)	DE (12)	TE (13)
CY (24)	SE (20)	BE (19)	LX (79)	WR (46)
DG (37)	ZK (21)	BL (78)	LY (21)	
GO (16)		BR (9)	SA (23)	

Vocalizations were recorded during the course of focused observational research on the baboons. Observational research involved consecutive 1-hour follows of specific (focal) animals and was conducted throughout the day but concentrated between 6:00–14:00 as the baboons left their sleeping site in the morning and then moved, foraged, and socialized for the first few hours each day. Behavioral data collected included a continuous record of the focal animal's general activity state (e.g., resting, moving, foraging), social behavior, vocalizations, and additional information related to spatial and ecological factors (e.g., group dispersion, location within the home range, habitat visibility). For additional details regarding the collection of behavioral data, see Rendall *et al.* (2000).

Special emphasis was placed on recording vocalizations produced by the focal animal and by other animals interacting with it. For each vocal recording, a detailed commentary was made of the contextual details associated with calling, including information on the identity of both the caller and others in the vicinity, on the caller's behavior immediately before and after calling (e.g., moving, feeding, socializing), on others' responses to the calls, and on the general ecological circumstance (e.g., habitat type, visibility). All recordings were made on Sony type IV metal tapes using a Sony WM-D6C Professional Walkman cassette recorder and a Sennheiser ME 80 directional microphone with K3U power module.

2. The grunt sample and acoustic analyses

From the database of all grunt recordings, we considered for inclusion in this study only those that were of the highest possible quality, uncontaminated by either environmental noise or the calls of other animals. We then selected grunt recordings for as many different adult males and females as possible. The sample ultimately available for acoustic analysis consisted of 1028 grunts from 27 different individuals (11 males and 16 females). Table I provides a more detailed description of the grunt sample.

Grunt recordings were then digitized for analysis with 16-bit accuracy using a sampling rate of 22.05 kHz, after low-pass filtering at 10 kHz. Acoustic analyses were performed using Version 5.3 of the ENTROPIC SIGNAL PROCESSING SYSTEM WAVES+ software package (ESPS/WAVES+©,

TABLE II. Sex differences in the acoustic features of grunts. Values shown are the grand mean for each acoustic feature. *F*-values are from ANOVA on individuals' means using sex as the factor (df 1, 25).

Acoustic feature (Hz)	Male ($N=11$)	Female ($N=16$)	*F*-value
F_0	53	114	*267.16*[a]
$F1$	332	448	*48.04*[a]
$F2$	1062	1430	*111.53*[a]
$F3$	1878	2677	*213.47*[a]

[a]$P<0.0001$.

1997) implemented on the Linux operating system (Red Hat 6.0©, 1999). Analyses focused on the extraction of mean values for the F_0 and the first three formant frequencies of each grunt.

Mean F_0 values were obtained from the amplitude waveform of each grunt. Because most grunts exhibited extremely regular periodicity, individual cycles of vocal-fold opening and closing (representing F_0) were clearly visible in the waveforms. We measured the period of each cycle of vocal-fold opening and closing from six cycles of vocal-fold vibration from the middle of each grunt. Most grunts contained 8–12 such cycles in total, so that the middle six cycles encompassed 50%–75% of each grunt around its midpoint. To obtain a mean F_0 value (in Hz) for each grunt, we calculated the mean of the inverse of each of the six period values for each call. The resulting mean F_0 value for each grunt was corroborated by comparing it against the first harmonic of a high-resolution (2048-point) frequency decomposition (fast Fourier transform, or FFT) of the same call taken from its midpoint.

The formant frequencies of each grunt were identified via autocorrelation-based linear predictive coding (LPC). In this analysis, an LPC spectrum was generated from a 1024-point (approximately 50 ms) segment from the midpoint of each grunt, using a Hanning window, 18 coefficients, and no pre-emphasis. This LPC spectrum was then overlaid on an independently derived FFT for the same segment of the grunt to ensure the goodness of fit of the LPC-derived frequency spectrum. The frequencies of the first three poles of the LPC spectrum were then extracted and retained. Preliminary screening and previous analyses (Owren *et al.*, 1997) showed that the formant frequencies of grunts do not vary appreciably within single calls, suggesting that dynamic modification of vocal-tract cavities during call production is uncommon. As a result, the formant structure of single grunts could be characterized accurately using a single, broad spectral slice from their midsection.

The resulting F_0 and formant values for each individual were then tested for potential sex differences using analysis of variance (ANOVA) implemented in the Number Cruncher Statistical System (NCSS, Hintze, 1999).

B. Results

The 16 adult females in the sample contributed a total of 747 grunts. Across all females, the mean F_0 of grunts was 114 Hz, while the means for the first three formant frequencies were 448, 1430, and 2677 Hz, respectively (Table II). The 11 adult males in this study contributed a total of 280 grunts. By comparison to adult females, the mean F_0 and formant frequencies of male grunts were appreciably lower (Table II). The mean F_0 of male grunts was 53 Hz, while the means of the first three formants were 332, 1062, and 1878 Hz. To test the statistical significance of these differences, separate ANOVAs with sex as the factor were conducted on each of the four acoustic parameters using as data each individual's mean value for that parameter. In each case, the difference between males and females was highly significant ($P<0.001$, Table II). Examples of these differences in the F_0 and formants of male and female grunts are shown graphically in Fig. 2.

Despite these clear sex differences in the mean F_0 and formants of grunts, there was sufficient variation in the frequency range of these features within and between individuals of both sexes to create some acoustic overlap between them. This point is illustrated in Fig. 3, which plots the mean (± 1 s.d.) F_0 and formant values for each individual in the sample. The figure clearly illustrates the general separation between males and females in each acoustic parameter, but with F_0 values showing the greatest effect. Formant frequencies were more likely to show some overlap, particularly $F1$, and to a lesser extent also $F2$.

III. PERCEPTUAL DISCRIMINATION OF SEX DIFFERENCES IN THE ACOUSTIC STRUCTURE OF GRUNTS

A. Methods

Experiments testing the perceptual discriminability of these sex differences in the acoustic structure of grunts involved captive baboons maintained in the Psychophysics Laboratory of the Johns Hopkins University School of Medicine. The laboratory baboons were tested on their ability to discriminate between the grunts of adult male and female baboons recorded in the wild by training them to perform in a reaction-time procedure in which the target stimuli (''release'' signals) were female baboon grunts, and the nontarget stimuli (''nonrelease'' signals) were male baboon grunts.

1. Experimental subjects

Five adult male baboons (*Papio hamadrayas anubis*, formerly *P. cynocephalus anubis*) were used as subjects. Each baboon was housed individually in a large-primate cage equipped with a seating bench, but had auditory and visual contact with other baboons housed in the same colony room. The animals were maintained on a 22 h restricted feeding schedule with water continuously available in the home cage. Supplemental monkey chow and two pieces of fresh fruit were provided daily after each experimental session. The baboons were maintained on a daily 12-h light/dark cycle (6 a.m./6 p.m.). Animal care was in accordance with current NIH guidelines on ''Principles of laboratory animal care.''

2. Perceptual testing apparatus

The perceptual testing apparatus involved a modified primate squeeze cage placed inside a double-walled, sound-attenuating chamber (IAC Inc., model 1201 A). An intelligence panel (76×97 cm) was attached to one side of the cage

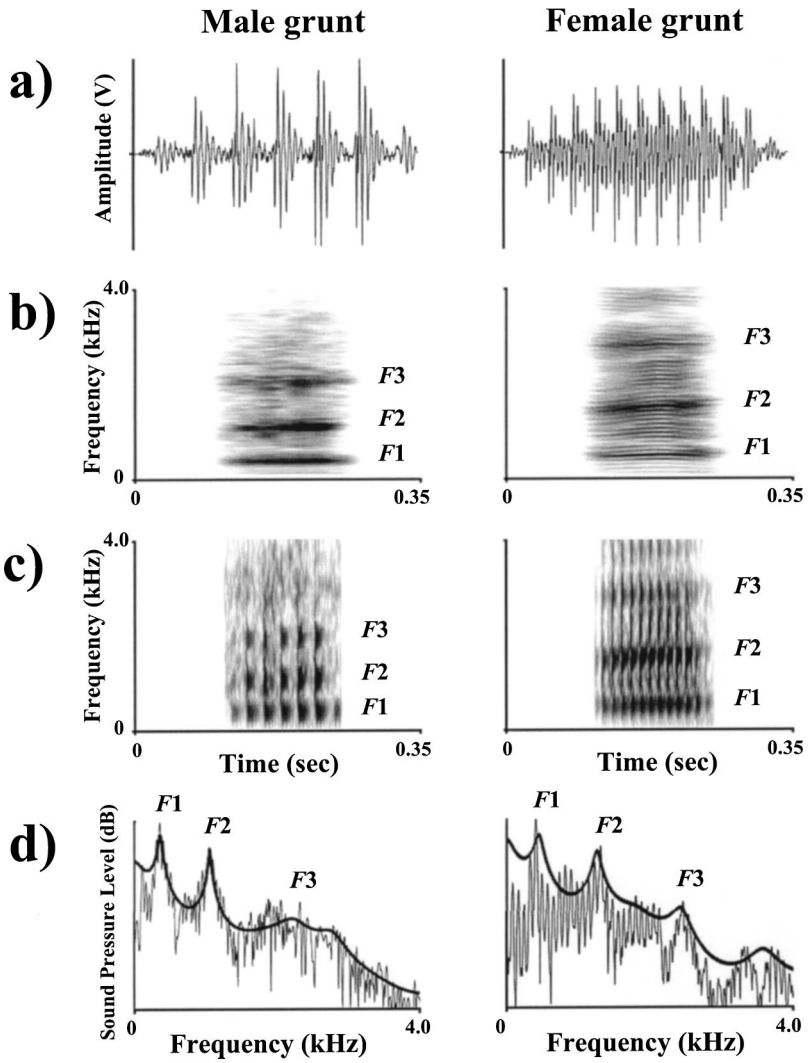


FIG. 2. Male and female baboon grunts shown (a) as time-series waveforms; (b) in narrow-band spectrograms (0.058-s Gaussian analysis window); and (c) in wideband spectrograms (0.0006-s Gaussian window). Panel (d) shows an LPC-based spectrum (14 coefficients, autocorrelation method) overlaid on a Fourier transform (Gaussian window), with each computed over the entire length of the respective calls. The first three formants visible in the spectrograms and LPC representations are marked as *F*1, *F*2, and *F*3.

and contained a red, light-emitting diode that served as a cue light, a feeder opening for delivery of 500-mg banana-flavored pellets, and a primate lever (BRS/LVE model PRL-003). With a baboon seated on a bench facing the panel, the cue light was at eye level, the feeder opening was 25 cm below the cue light, and the response lever was at waist level in front of the animal's right arm. Each baboon was moved from his home cage to the test cage via a metal transfer cage.

Digital audio equipment from Tucker-Davis Technologies (TDT) was used to control and present grunt tokens to the baboons. Signals were sampled at 20 kHz, passed through a programmable attenuator, an amplifier, and then to a wide-range speaker in each test chamber. The stimuli were presented at an average level of 75 dB. To prevent subjects from responding to possible differences in intensity among stimuli, amplitude levels of all stimuli were varied by randomly selecting an intensity prior to each stimulus presentation. Stimulus intensity was varied over a ±3-dB range in 1-dB steps. Stimulus presentations, response measures, and contingencies were controlled by computers.

3. Discrimination training

The start of each trial in the reaction-time procedure was signaled by a flashing red cue light (5/s). A lever press changed this flashing light to a continuous light, which remained steady as feedback as long as an animal held the lever down. At intervals ranging from 1.0 to 7.0 s after initiation of this maintained holding response, a signal was presented for 1.5 s (''response window''). If the signal was a female grunt, release of the lever within the response window (a ''hit'' response) delivered the reinforcer (one 500-mg banana-flavored pellet); if the signal was a male grunt, release of the lever (a ''false alarm'') produced a 5–10 s timeout. In both cases, lever release also initiated a 3-s intertrial interval (ITI) during which no stimuli were presented and additional lever responses reinitiated the ITI. Lever releases prior to any grunt produced a 5–10-s timeout without reinforcement. If an animal failed to release the lever following a female grunt (a ''miss''), the red cue light was turned off at the end of the response window to indicate the end of the trial, and subsequent lever release then returned the animal to the ITI. If an animal failed to release the lever within the response window following a male grunt (a ''correct rejection''), a reinforcer was delivered followed by the red cue light being turned off and subsequent lever release then returning the animal to the ITI. Following the end of the ITI, the cue light began flashing to signal initiation of the next trial in the series of several hundred which comprised each daily session.

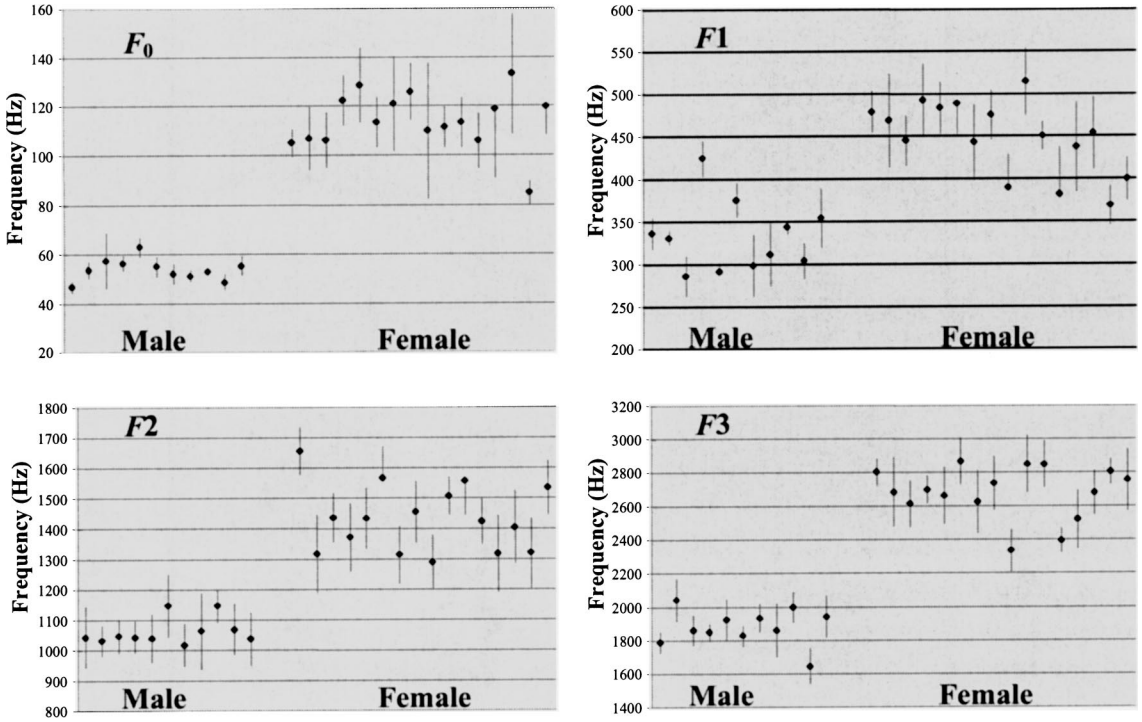


FIG. 3. Mean frequency values (±s.d.) of the F_0 and first three formants of grunts for individual adult males and females.

Sessions were 100 min in duration and occurred 5 days a week at approximately the same time each day. During a session, baboons typically performed about 500 discrimination trials. For each comparison grunt, the percent-correct score was defined as the number of releases within the 1.5-s response window divided by the total number of trials presented, multiplied by 100. Baseline performances were defined as stable when the percentage of correct responses to all female grunts was 80% or greater during a session, false alarm rates to the male grunts was less than 20% during a session, and there were no systematic changes in the time course of these measures within or across sessions.

4. Stimulus conditions during training and testing

All baboons were initially trained with a single male grunt (male J; see Table III) as the background stimulus and a single female grunt (female A) as the target stimulus. Once they were reliably releasing the lever to the female grunt but holding the lever to the male grunt, a comprehensive series of experimental tests was undertaken to examine the ability of the subjects to discriminate between the grunts of different adult males and females. The six experimental conditions are outlined in Table III and were designed to determine whether the baboons would generalize their ''releasing response'' to (1) new grunt tokens from the original female; (2) grunts from different (novel) females; (3) additional grunt tokens from these novel females; and (4) grunts from yet another set of completely different females. Conversely, experimental conditions were designed to determine whether subjects would also generalize their ''holding response'' to (5) a different grunt from the original male; and (6) grunts from completely different (novel) males.

All grunts used in these experiments derived from field recordings of wild baboons. As a result, the individuals rep-

TABLE III. Experimental conditions and call exemplars used in behavioral training and testing. Letters A–H refer to individual females, and J–N refer to individual males, whose calls were used in the different conditions; #'s denote different call tokens from these individuals.

Experimental condition	Male standard grunt	Female target grunts			
		Grunt #1	Grunt #2	Grunt #3	Grunt #4
Initial training	J	A	A	A	A
(1) Original female, new tokens	J	A	A2	A3	A4
(2) Novel females	J	A	B	C	D
(3) Novel females, new tokens	J	A5	B2	C2	D2
(4) Additional novel females	J	E	F	G	H
(5) Original male, new token	J2	A	B	C	D
(6) Novel males	K	A	B	C	D
	L	A	B	C	D
	M	A	B	C	D
	N	A	B	C	D

resented by the grunt stimuli were unfamiliar to the subjects. In this way, we could ensure that subjects' discrimination of grunts in these experiments was not based on known characteristics of familiar individuals' voices.

5. Behavioral data analysis

To evaluate subjects' discrimination performance, we used the standard signal detection index, d' (Macmillan and Creelman, 1991). In addition to characterizing subjects' correct discriminations, this index controls for performance variations between subjects. The d' index was calculated by transforming the percent-correct scores (PC) and false alarm rates (FA) into proportions. Each of these proportions was then converted into the equivalent of a z score by calculating the inverse of the standard normal cumulative distribution for each proportion. The d' index was then calculated by subtracting the FA z scores from the PC z scores, i.e., $d' = z(\mathrm{PC}) - z(\mathrm{FA})$.

To estimate each baboon's initial reactions to the stimuli presented in each of the six experimental conditions, d' index values were calculated only for the first 20 trials of exposure to each stimulus, and only for data obtained in the *first* session in which new stimuli were presented. For comparison purposes, ''control'' data were similarly based upon the first 20 trials of exposure to each stimulus, but were taken from the *last* session that each baboon performed under the immediately preceding testing condition. The control data were subsequently averaged across the six experimental conditions listed in Table III, and thus represent the average performance immediately prior to each of the experimental testing conditions.

B. Results

Figure 4 summarizes the mean discrimination performance of the five subjects in the four experimental conditions involving different female calls (i.e., conditions 1, 2, 3, and 4 from Table III). Each bar in these graphs represents the overall mean ($\pm$SEM) d' index value for a given stimulus comparison (i.e., the mean of individuals' mean d' index values for that comparison). The dotted line corresponds to a d' index value of 1.0, an accepted standard for reliable discriminations.

Figure 4(a) shows that across subjects the mean d' index value exceeded 2.0 on the first 20 trials of the first session in which three new calls from the original female used in baseline training were introduced (experimental condition 1). Thus, on first exposure to new call tokens of the original female, subjects accurately discriminated them from the male standard, indicating immediate and complete generalization to new grunts of the same female. Figure 4(b) shows that when calls of novel females were introduced (experimental condition 2), discrimination performances remained high, with mean d' index values between 2.0 and 3.0. Thus, subjects immediately discriminated these calls from the male standard, indicating that they also readily generalized to grunts from other females. Figure 4(c) shows that when different tokens of these new females were added (experimental condition 3) discrimination performance was unaffected.

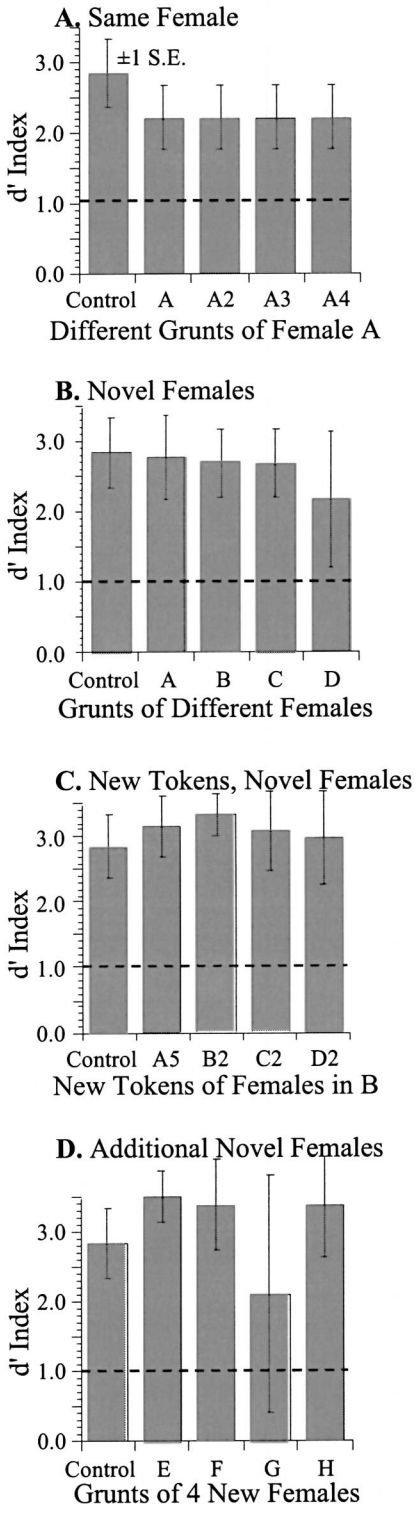


FIG. 4. Subjects' discrimination performance (d') in experimental conditions using different female calls. Panels A, B, C, and D correspond, respectively, to conditions 1, 2, 3, and 4 from Table III. Each bar represents the overall mean ($\pm$SEM) d' value across subjects for a particular stimulus contrast based on the first 20 trials of the first session in which each new stimulus was presented.

Mean d' index values remained uniformly high at a level of 3.0 or more. Finally, Fig. 4(d) shows that when calls from yet another set of novel females were added (experimental condition 4), discrimination performance remained high with a single possible exception. For one of the four new female

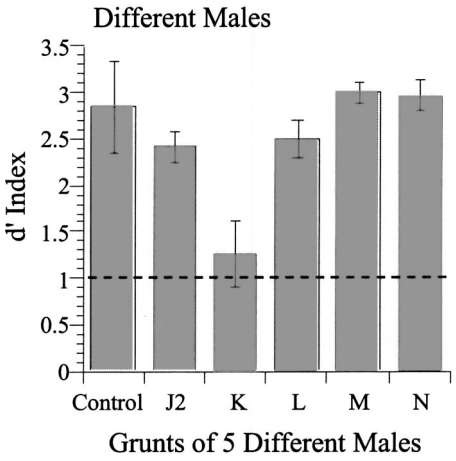


FIG. 5. Subjects' discrimination performance (d') in experimental conditions using different male calls (conditions 5 and 6 from Table III). Other details are as per Fig. 4.

grunts tested in this condition (G), the mean d' index value across subjects was approximately 2.0, a level of discrimination still considerably higher than 1.0. However, there was also considerable variation in discrimination performance across subjects for this one call. Four of the five subjects showed little or no decrement in performance, with d' index values of approximately 3.0, while performance for one subject dropped to a d' index value just below 1.0, suggesting that there was something about this particular call that made it difficult to discriminate from the standard male for this one subject. However, all five subjects readily discriminated the calls of the remaining three females tested in this condition, with mean d' index values exceeding 3.0. This outcome confirmed their earlier ready discrimination of novel females from the standard male in condition 2.

Figure 5 summarizes the mean discrimination performance of the five subjects in the two experimental conditions involving different male calls (i.e., conditions 5 and 6 from Table III). When the standard male grunt (J) was changed to a different token of the same male (J2, experimental condition 5), subjects' discrimination of the females' grunts from this new male grunt was virtually unchanged. The mean d' index value remained at approximately 2.5, the same level as in initial discrimination training. Thus, subjects immediately generalized their discrimination to a different grunt of the same male used in baseline training. When grunts of novel males were subsequently introduced (experimental condition 6), subjects' performance declined at first but then rebounded immediately to baseline levels. Thus, with the addition of the first novel male (male K), mean d' index values fell to approximately 1.25. Here again, the low overall performance was attributable primarily to the results for two subjects whose mean d' index values dropped below 1.0. The other subjects showed little decrement in performance with d' index values of approximately 2.0. With the addition of three additional males (L,M,N), discrimination performance for all five subjects increased immediately, with mean d' index values between 2.5 and 3.0.

In summary, 20 different stimulus comparisons were undertaken to test the ability of five subjects to discriminate grunts produced by adult males and females. In only two of the 20 comparisons did the performance of any of the subjects fall to a d' index value close to 1.0, and in both cases it was the performance of only one or two subjects that was affected. In the remaining 18 comparisons, the discrimination performance of all subjects was uniformly high, with d' index values equal to or greater than levels attained during baseline discrimination training (i.e., $d' \geqslant 2$). Thus, performance levels were considerably higher than the threshold for reliable discrimination in 97 of the 100 total subject-by-stimulus tests. Overall, then, subjects appeared readily able to discriminate the grunts of adult males and females, showing clear and immediate generalization of their accurate discrimination of grunts from the standard male and female in baseline training not only to additional grunts from these two individuals but also to grunts from completely novel males and females.

IV. DISCUSSION

Analyses of the grunts of adult male and female baboons revealed clear sex differences in the mean F_0 and the means of each of the first three formant frequencies. For both the F_0 and formants, mean frequency values for adult males were lower than they were for adult females (Table II). On average, the F_0 of male grunts was approximately 50% of that of female grunts, while the formant frequencies of male grunts were 70%–75% of those of female grunts. The latter finding is consistent with the results of a recent study of the long-distance, loud calls (or "barks") produced by male and female chacma baboons (Fischer *et al.*, 2002). In that study, the mean frequency of each of the first two formants (or "dominant frequency bands") in male barks was approximately 77% of that in females. Because barks tend to be harsher (less tonal) than grunts, acoustic analyses of them could not include a measure of F_0.

The pattern of sex differences in F_0 and formants revealed here for baboon grunts also closely parallels the situation in human speech, where mean frequency values for both acoustic features are also lower in males than they are in females and more dramatically so for F_0 than for formants. Although absolute values vary marginally between studies depending on the details of the sample used and the range of vowel sounds studied (e.g., Peterson and Barney, 1952; Baken, 1987; Hillenbrand *et al.*, 1995; Stevens, 1998), the mean F_0 of relaxed speech in human males is approximately 120 Hz, while that in females is approximately 220 Hz, and mean frequencies of the first three formants in males are approximately 500, 1500, and 2500 Hz, respectively, while those in females are approximately 600, 1750, and 2850 Hz. Thus, the mean F_0 in male speech is approximately 55% of that in females, while mean formant values in males are approximately 85% of those in females.

Results of our laboratory perceptual testing suggest that the acoustic–structural differences in the grunts of male and female baboons are also discriminable to the animals themselves. After an initial training period to learn to discriminate the grunt of one male from that of one female, all baboons were subsequently able to discriminate new grunt tokens from these two individuals as well as additional grunts from

completely novel males and females. Because the callers were unfamiliar to subjects, these discriminations could not have been based on known characteristics of the voices of individual callers, but rather must have been based on acoustic–structural differences in the calls associated with the callers' different sex.

Of course, the ability of the baboons to correctly discriminate between the grunts of males and females is not surprising given that the magnitude of the acoustic–structural differences between their grunts is likely to be well within the species' frequency discrimination abilities. Although frequency difference limens have not yet been published for baboons, those of closely related macaque monkeys are known to be approximately 10–60 Hz for pure-tone stimuli below 2 kHz, and approximately 30–50 Hz for formant-type stimuli in the same frequency range (Sinnott and Kreiter, 1991; Sommers *et al.*, 1992). Research in progress indicates that the discrimination profile for baboons is very similar to that for macaques for both pure-tone and formant-type stimuli (Hienz, unpublished data). Hence, the sex difference in mean F_0 (approximately 60 Hz) and mean formant frequencies (ranging from approximately 120 Hz for $F1$ to 600 Hz for $F3$), are likely to be well above the corresponding frequency discrimination thresholds in baboons.

Parallel work on humans likewise shows ready discrimination of the voices of unfamiliar male and female talkers (e.g., Whiteside 1998a). Here again, accurate discrimination of speaker sex is not surprising as the marked differences in F_0 and formants in male and female speech also far exceed frequency discrimination thresholds in the species (Bachorowski and Owren, 1999). Reported frequency difference limens in humans for tonal- and formant-type stimuli are approximately 1–2 and 25 Hz, respectively (Wier *et al.*, 1977; Kewley-Port and Watson, 1994), while the difference in mean F_0 in humans is approximately 100 Hz, and the differences in mean formant frequencies range from 100 Hz (for $F1$) to 350 Hz (for $F3$).

V. IMPLICATIONS FOR FUTURE WORK

We have described basic patterns of sex difference in the acoustic structure of baboon grunts and documented the perceptual discrimination of these differences by baboon listeners. In both cases, the results parallel equivalent phenomena in human speech production and perception. The next important research steps, then, are to examine whether the underlying anatomical bases for the acoustic–structural differences in the vocal signals of the two species are the same, and whether the perceptual processes involved in discriminations of vocalizer sex in each species are also similar.

Sex differences in the mean F_0 and formants of vocal signals in both species probably arise in large part from sex differences in the length and mass of the vocal folds and the length and diameter of the supralaryngeal vocal-tract cavities, respectively. These vocal anatomical differences in turn may be traceable ultimately to sex differences in body size (Fitch, 1997; Fitch and Giedd, 1999; Fischer *et al.*, 2002). We are not aware of any systematic morphometric data on vocal-fold or vocal-tract length in chacma baboons that might corroborate the direct vocal anatomical basis for the sex difference in F_0 and formants in this species. However, there is a reasonable corpus of data on basic body size from several wild populations. Most of the data pertain to body weight, rather than measures of body height or length, but indicate that adult males are approximately twice the size of adult females, on average weighing about 30 kg compared to females' 15 kg (reviewed in Barrett and Henzi, 1997). Thus, the twofold difference in the mean F_0 of grunts produced by adult males and females matches the twofold difference in body size between them. However, this body-size difference is considerably greater than the 25% difference in formant frequencies observed between the sexes.

Interestingly, this pattern of asymmetric correspondence between acoustic–structural features and overall body size appears to be reversed in humans. On average, adult males are about 10% taller and 20% heavier than adult females (Alexander *et al.*, 1979; Hamill *et al.*, 1979; Rendall and Owren, 2003). In contrast to the situation in baboons, then, the almost twofold difference in mean F_0 between the sexes in humans greatly exceeds their more modest difference in overall body size, whereas the approximately 15% difference in formants corresponds comparatively well to the body-size difference. Available morphometric data on human vocal anatomy help to clarify these relationships. They show that, by comparison to adult females, the supralaryngeal vocal tract in adult males is approximately 10%–15% longer (Fant, 1966; Fitch and Giedd, 1999) and the vibrating segment of the vocal folds is roughly 60% longer (Hirano, 1983; Titze, 1989, 1994). In other words, in humans, the sex difference in formants appears to stem from a difference in vocal-tract length between males and females, and this in turn corresponds relatively closely to the overall body-size difference between the sexes. In contrast, although the sex difference in F_0 also seems to match pretty closely the difference in vocal-fold length between males and females, this latter difference appears to reflect either a dramatic exaggeration in males, or a dramatic reduction in females, of vocal-fold dimensions relative to overall body size (for possible functional evolutionary explanations of these phenomena see Fitch and Giedd, 1999).

Thus, although the absolute sex differences in F_0 and formants in baboons and humans appear to be very similar, the connection of these acoustic–structural differences to underlying vocal anatomical and overall body-size differences in each species is less clear. In humans, formant differences between the sexes appear to track differences in overall body size between them relatively well, while F_0 differences appear to reflect dramatic changes in vocal-fold dimensions relative to body size. In contrast, in baboons, F_0 differences between the sexes appear to match body-size differences between them, but formant differences do not. It is possible that the formant mismatch in baboons is illusory, that formant differences actually track body-size differences between the sexes, only in linear as opposed to volumetric body dimensions. Measures of body height, or length, have not typically been available from field studies of baboons. However, if body weight scales as approximately the cube of linear body dimensions (as it should), then the twofold difference in body weight observed in baboons should represent an only

approximately 25% difference in linear dimensions (e.g., body length or height). If true, then the roughly 25% difference in formants between adult male and female baboons observed here would, in fact, closely match the difference in linear body size between them. Convincingly resolving this issue will require more systematic data on both linear and volumetric body dimensions in baboons. These data are currently being collected.

Whatever the ultimate anatomical bases prove to be for the sex differences in acoustic–structural features of baboon grunts and human speech, the differences themselves represent a potentially important source of acoustic variability that listeners may have to normalize in order to interpret the functional significance of the signals. This is certainly true for human speech, where listeners must somehow discount the baseline formant differences in the speech of males and females in order to distill a ''normalized'' formant pattern that conveys phoneme identity. Some similar process may operate in baboons whose grunts appear to involve different functional variants differing in features related to vocal-tract resonance (Owren *et al.*, 1997; Rendall *et al.*, 1999).

In the present study, we have shown that the baboons can discriminate between the grunts of males and females. The next step is therefore to examine the basis of this discrimination in more detail, specifically by testing the relative perceptual importance of F_0 and the various formants. Furthermore, the methodology of the current work does not allow a strong conclusion that the baboons were responding to the calls as members of discrete classes, namely male versus female grunts. That interpretation seems likely, as a male–female vocal difference is a basic and natural distinction that is probably a salient component of baboon social interactions across a variety of contexts. However, a more involved testing procedure will be required to definitively demonstrate that baboon listeners can show truly categorical sorting of grunts by sex, for instance through a two-alternative, forced-choice design that requires explicit and symmetrical classification on each trial.

It will also be important to test the extent to which discriminations of caller sex involve disproportionate weighting of the acoustic–structural differences in either F_0 or formant frequencies. Given the more dramatic relative sex difference in F_0 compared to formant frequencies, F_0 might be expected to take precedence here. However, F_0 also varies markedly with caller motivation and arousal in baboons (Rendall, 2003), suggesting that it may not be a perfectly reliable source for discriminating caller sex. Furthermore, because functionally distinct variants of grunts appear to differ primarily in spectral resonance patterns (Owren *et al.*, 1997; Rendall *et al.*, 1999), the process of normalizing these differences to accurately identify particular grunt variants might make baboons especially sensitive to the formant differences between sexes, although it is difficult to predict which specific formant(s) might drive such a bias.

The very same set of issues also applies in humans, of course, and has received some direct empirical testing although with as yet equivocal results. Thus, there is some evidence that, consistent with its more dramatic relative frequency difference also in humans, F_0 takes precedence over formants in discriminations of speaker sex (Childers and Wu, 1991; Whiteside, 1998b), even though, as in baboons again, F_0 can vary markedly according to the speaker's motivational and arousal state and other paralinguistic dimensions of speech. At the same time, though, there is some evidence indicating that, consistent with its comparative linguistic prominence, formant variation is also important in discriminating speaker sex (Coleman, 1976; Whiteside, 1998b), although to our knowledge there is none that addresses the relative influence of any specific formant(s) on this effect.

A clear resolution of the relative importance of F_0 and formants in discriminations of vocalizer sex, and by extension also in potential vocalizer normalization processes, awaits additional systematic experimentation, certainly in baboons and perhaps also still in humans. However, given the apparent parallels between the two species both in the basic acoustic structural features of their respective communicative signals and in the various constitutional and communicative dimensions they reflect (e.g., sex, size, motivation, and ''linguistic'' functions), the work on baboons and humans should ultimately prove to be reciprocally informative.

ACKNOWLEDGMENTS

We are grateful to the Office of the President and the Department of Wildlife and National Parks of the Republic of Botswana for permission to conduct research in the Moremi Game Reserve, and to the Department of Home Affairs and the Cape Nature Conservancy of South Africa for permission to conduct research in the De Hoop Nature Reserve. We also thank R. Seyfarth and D. Cheney, and P. Henzi and L. Barrett for the opportunity to work with baboons in Botswana and South Africa, respectively. Thanks also to L. Christensen, C. Pyle, and A. Jones for their assistance with research. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada, the L.S.B. Leakey Foundation, and the University of Lethbridge (DR), and by Grant No. DA 12139 from the National Institute on Drug Abuse, U.S.A. (RDH).

Alexander, R. D., Hoogland, J. L., Howard, R. D., Noonan, K. M., and Sherman, P. W. (**1979**). ''Sexual dimorphism and breeding systems in pinnipeds, ungulates, primates, and humans,'' in *Evolutionary Biology and Human Social Behavior: An Anthropological Perspective*, edited by N. A. Chagnon and W. Irons (Duxbury, North Sciuate, MA), pp. 402–455.

Andrew, R. J. (**1976**). ''Use of formants in the grunts of baboons and other nonhuman primates,'' Ann. N.Y. Acad. Sci. **280**, 673–693.

Bachorowski, J.-A., and Owren, M. J. (**1999**). ''Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech,'' J. Acoust. Soc. Am. **106**, 1054–1063.

Baken, R. J. (**1987**). *Clinical Measurement of Speech and Voice* (Allyn and Bacon, Toronto).

Barrett, L., and Henzi, S. P. (**1997**). ''An inter-population comparison of body weight in chacma baboons,'' S Afr. J. Sci. **93**, 436–438.

Childers, D. G., and Wu., K. (**1991**). ''Gender recognition from speech. II. Fine analysis,'' J. Acoust. Soc. Am. **90**, 1841–1856.

Coleman, R. O. (**1976**). ''A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness in the voice,'' J. Speech Hear. Res. **19**, 168–180.

Collins, S. A. (**2000**). ''Men's voices and women's choices,'' Anim. Behav. **60**, 773–780.

ESPS/waves+© (**1997**). *Entropic Research Laboratory, Inc.*, 600 Pennsylvania Ave. S.E. Suite 202, Washington, D.C., 2003.

Fant, G. (**1966**). ''A note on vocal tract size factors and non-uniform F-pattern scaling,'' Speech Transactions Laboratory Quarterly Progress and Status Report **4**, 22–30.
Fischer, J., Hammerschmidt, K., Cheney, D. L., and Seyfarth, R. M. (**2002**). ''Acoustic features of male baboon loud calls: Influences of context, age, and individuality,'' J. Acoust. Soc. Am. **111**, 1465–1474.
Fitch, W. T. (**1997**). ''Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques,'' J. Acoust. Soc. Am. **102**, 1213–1222.
Fitch, W. T. (**2000**). ''The evolution of speech: A comparative review,'' Trends Cog. Sci. **4**, 258–267.
Fitch, W. T., and Hauser, M. D. (**1995**). ''Vocal production in nonhuman primates: Acoustics, physiology, and functional constraints on honest advertisement,'' Am. J. Primatol **37**, 179–190.
Fitch, W. T., and Giedd, J. (**1999**). ''Morphology and development of the human vocal tract. A study using magnetic resonance imaging,'' J. Acoust. Soc. Am. **106**, 1511–1522.
Hamill, P. V. V., Drizd, T. A., Johnson, C. L., Reed, R. B., Roche, A. F., and Moore, W. M. (**1979**). ''Physical growth: National Center for Health Statistics percentiles,'' Am. J. Clin. Nutr. **32**, 607–629.
Hauser, M. D. (**1996**). *The Evolution of Communication* (MIT Press, Cambridge, MA).
Hienz, R. D., and Brady, J. V. (**1988**). ''The acquisition of vowel discriminations by nonhuman primates,'' J. Acoust. Soc. Am. **84**, 186–194.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.
Hintze, J. L. (**1999**). *Number Cruncher Statistical System, 2000* (Kaysville, Utah).
Hienz, R. D. (unpublished data).
Hirano, M., Kurita, S., and Nakashima, T. (**1983**). ''Growth, development, and aging of human vocal folds,'' in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. M. Bless and J. H. Abbs (College-Hill, San Diego), pp. 22–43.
Johnson, K., and Mullenix, J. W. (**1997**). *Talker Variability in Speech Processing* (Academic, New York).
Kewley-Port, D., and Watson, C. S. (**1994**). ''Formant-frequency discrimination for isolated English vowels,'' J. Acoust. Soc. Am. **95**, 485–496.
Kunzel, H. J. (**1989**). ''How well does average fundamental frequency correlate with speaker height and weight,'' Phonetica **46**, 117–125.
Ladefoged, P., and Broadbent, D. E. (**1957**). ''Information conveyed by vowels,'' J. Acoust. Soc. Am. **39**, 98–104.
Lass, N. J., and Brown, W. S. (**1978**). ''Correlational study of speakers' heights, weights, body surface areas, and speaking fundamental frequencies,'' J. Acoust. Soc. Am. **63**, 1218–1220.
Lieberman, P. (**1968**). ''Primate vocalizations and human linguistic ability,'' J. Acoust. Soc. Am. **44**, 1574–1584.
Lieberman, P. (**1986**). ''Some aspects of dimorphism and human speech,'' in *Sexual Dimorphism in Living and Fossil Primates*, edited by M. Pickford and B. Chiarelli (Il Sedicesimo, Italy), pp. 67–75.
Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge).
Nearey, T. (**1989**). ''Static, dynamic, and relational properties in vowel perception,'' J. Acoust. Soc. Am. **85**, 2088–2113.
Negus, V. E. (**1949**). *The Comparative Anatomy and Physiology of the Larynx* (Hafner, New York).
Owren, M. J., and Linker, C. D. (**1995**). ''Some analysis methods that may be useful to acoustic primatologists,'' in *Current Topics in Primate Vocal Communication*, edited by E. Zimmermann, J. D. Newman, and U. Jurgens (Plenum, New York), pp. 1–27.
Owren, M. J., Seyfarth, R. M., and Cheney, D. L. (**1997**). ''The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cynocephalus ursinus*): Implications for production processes and functions,'' J. Acoust. Soc. Am. **101**, 2951–2963.
Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24**, 175–184.
Red Hat LINUX, Version 6.0© (**1999**). Red Hat Software (Durham, NC).
Rendall, D. (**2003**). ''Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons,'' J. Acoust. Soc. Am. **113**, 3390–3402.
Rendall, D., and Owren, M. J. (**2003**). ''The sounds of sex: Are they all simply a matter of size?'' Paper presented at the Acoustical Society of America, International Conference on Animal Communication, University of Maryland (July 2003).
Rendall, D., Seyfarth, R. M., Cheney, D. L., and Owren, M. J. (**1999**). ''The meaning and function of grunt variants in baboons,'' Anim. Behav. **57**, 583–592.
Rendall, D., Cheney, D. L., and Seyfarth, R. M. (**2000**). ''Proximate factors mediating 'contact' calls in adult female baboons and their infants,'' J. Comp. Psychol. **114**, 36–46.
Richman, B. (**1976**). ''Some vocal distinctive features used by gelada monkeys,'' J. Acoust. Soc. Am. **60**, 718–724.
Sinnott, J. M., and Kreiter, N. A. (**1991**). ''Differential sensitivity to vowel continua in Old World monkeys (Macaca) and humans,'' J. Acoust. Soc. Am. **89**, 2421–2429.
Sommers, M. S., Moody, D. B., and Prosen, C. A. (**1992**). ''Formant frequency discrimination by Japanese macaques (*Macaca fuscata*),'' J. Acoust. Soc. Am. **91**, 3499–3510.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).
Titze, I. R. (**1989**). ''Physiologic and acoustic differences between male and female voices,'' J. Acoust. Soc. Am. **85**, 1699–1707.
Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).
Trout, J. D. (**2001**). ''The biological basis of speech: What to infer from talking to the animals,'' Psychol. Rev. **108**, 523–549.
van Dommelen, W. A. (**1993**). ''Speaker height and weight identification: A re-evaluation of some old data,'' J. Phonetics **21**, 337–341.
Whiteside, S. P. (**1998a**). ''Identification of a speaker's sex: A study of vowels,'' Percept. Mot. Skills **86**, 579–584.
Whiteside, S. P. (**1998b**). ''Identification of a speaker's sex from synthesized vowels,'' Percept. Mot. Skills **86**, 595–600.
Wier, C., Jesteadt, W., and Green, D. M. (**1977**). ''Frequency discrimination as a function of frequency and sensation level,'' J. Acoust. Soc. Am. **61**, 178–184.
Zhinkin, N. I. (**1963**). ''An application of the theory of algorithms to the study of animal speech,'' in *Acoustic Behaviour of Animals*, edited by R. G. Busnel (Elsevier, Amsterdam), pp. 132–180.

Determination of West Indian manatee vocalization levels and rate

Richard Phillips and Christopher Niezrecki[a)]
Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, Florida 32611-6250

Diedrich O. Beusse
College of Veterinary Medicine, University of Florida, P.O. Box 100126, Gainesville, Florida 32610-0126

(Received 24 July 2003; revised 29 October 2003; accepted 3 November 2003)

The West Indian manatee (*Trichechus manatus latirostris*) has become endangered partly because of a growing number of collisions with boats. A system to warn boaters of the presence of manatees, based upon the vocalizations of manatees, could potentially reduce these boat collisions. The feasibility of this warning system would depend mainly upon two factors: the rate at which manatees vocalize and the distance in which the manatees can be detected. The research presented in this paper verifies that the average vocalization rate of the West Indian manatee is approximately one to two times per 5-min period. Several different manatee vocalization recordings were broadcast to the manatees and their response was observed. It was found that during the broadcast periods, the vocalization rates for the manatees increased substantially when compared with the average vocalization rates during nonbroadcast periods. An array of four hydrophones was used while recording the manatees. This allowed for position estimation techniques to be used to determine the location of the vocalizing manatee. Knowing the position of the manatee, the source level was determined and it was found that the mean source level of the manatee vocalizations is approximately 112 dB (*re* 1 μPa) @ 1 m. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1635839]

PACS numbers: 43.80.Ka, 43.80.Nd [WA] Pages: 422–428

I. INTRODUCTION

Between 1995 and 2002 the percentage of mortalities of the West Indian manatee (*Trichechus manatus latirostris*) due to watercraft strikes has risen from 22% to 31% (Florida Department of Environmental Protection, 1996; Florida Fish and Wildlife Conservation Commission, 2002). This has led to increased research into manatee avoidance technologies. One type of manatee avoidance technology is based on detecting the presence of manatees with hydrophones listening for manatee vocalizations. The frequency of manatee vocalizations, the source level of the vocalizations, and the volume of the ambient background noise would all affect the feasibility of an acoustic detection system. The characteristics of manatee vocalizations would also need to be known. A comprehensive literature review on manatee vocalizations can be found in the work by Niezrecki *et al.* (2003). Typical manatee vocalizations have a duration between 0.25 and 0.5 s. The vocalizations usually contain a fundamental frequency and several higher harmonics. For a majority of the manatee calls, the first harmonic past the fundamental frequency dominates and contains a majority of the acoustic energy. The fundamental frequency ranges from 2 to 5 kHz and tends to be smaller than the first harmonic. The magnitude of the higher harmonics decreases as the frequency increases. The vocalizations can be variable and for some calls the fundamental frequency dominates over the first harmonic.

Researchers have studied the rate at which manatee calls are made. Other than when they are feeding, manatees vocalize on average approximately one to five times within a 5-min period, depending on their activity (Bengston and Fitzgerald, 1985). Bengston and Fitzgerald observed manatee vocalization rates and the corresponding manatee behaviors. It was determined that vocalization rates are dependent on a manatee's behavior, with feeding and resting having lowest vocalization rates and mating and cavorting being the highest. It is also suggested that if manatee vocalizations are used for communicative and social purposes, then the vocalization rates may depend upon the number of manatees present. Nowacek compared the vocalization rates in Florida and Belize and found that the manatees vocalized at a rate of 1.29 and 0.09 to 0.75 vocalizations per animal per minute, respectively (Nowacek *et al.*, 2003). The high rate of vocalization at Crystal River is based on visual estimates of the number of manatees ($\sim$50) in the area. However, the visual estimate of the number of manatees is not an accurate assessment of the number of manatees in the river. Additionally, the behavior of the manatees during the measurement was not reported in their work and can affect the vocalization rate. Nowacek also estimated the mean received sound pressure levels of the peak frequency (in Crystal River) to be approximately 100 dB (*re* 1 μPa). The received values are recorded with the hydrophone at approximately 20 m from a group of 50 manatees. Additionally, the source levels from the hydrophones tagged on the Belize manatees were estimated to be between 106 and 115 dB.

This work contributes to the scientific knowledge of

[a)]Author to whom correspondence should be addressed. Electronic mail: niezreck@ufl.edu

manatees by confirming the manatee vocalizations rates determined by prior researchers. More importantly, this is the first work to measure the source levels of manatee vocalizations by using triangulation techniques and to document the response of manatees to broadcasts of manatee vocalizations.

II. THEORETICAL DEVELOPMENT

In order to calculate the source levels, the location of the source and the sensors must be determined. Two different position estimation methods are used to determine the location of the vocalizing manatee. The first method used is based upon the differences in the time-of-flight of the signal and the equation of a circle. The second method also uses the differences in the time-of-flight, but uses the equation of a hyperboloid. Both methods are now reviewed.

A. Time-of-flight differences

One of the position estimation methods used in this experiment was developed by Mahajan and Walworth (Mahajan and Walworth, 2001). This method uses the time-of-flight differences from four sensors to locate a source on a plane. The signal arrival times (T_j) at the four sensors are used to calculate the three time-of-flight differences (ΔT_{1j}) between sensor 1 and the remaining three sensors, as shown in Eq. (1):

$$\begin{aligned}\Delta T_{12}&=T_2-T_1,\\ \Delta T_{13}&=T_3-T_1,\\ \Delta T_{14}&=T_4-T_1.\end{aligned} \tag{1}$$

The four sensors are located at measured positions (x_1,y_1) through (x_4,y_4) and the unknown location of the source is (u,v). The distance between sensor 1 and the source location is represented by d. The distance between the remaining sensors and the source is then equal to $d+c\Delta T_{1j}$, where c is the speed of sound. Equation (2) is derived by using the equation of a circle four times around the source, one circle through each sensor:

$$\begin{aligned}(x_1-u)^2+(y_1-v)^2&=d^2,\\ (x_2-u)^2+(y_2-v)^2&=(d+c\Delta T_{12})^2,\\ (x_3-u)^2+(y_3-v)^2&=(d+c\Delta T_{13})^2,\\ (x_4-u)^2+(y_4-v)^2&=(d+c\Delta T_{14})^2.\end{aligned} \tag{2}$$

The first equation of a circle in Eq. (2) is expanded and substituted into the remaining three equations in place of d^2. The three resulting equations are then represented in Eq. (3), in matrix form:

$$\begin{bmatrix}2x_1-2x_2 & 2y_1-2y_2 & -2c\Delta T_{12}^2\\ 2x_1-2x_3 & 2y_1-2y_3 & -2c\Delta T_{13}^2\\ 2x_1-2x_4 & 2y_1-2y_4 & -2c\Delta T_{14}^2\end{bmatrix}\begin{bmatrix}u\\ v\\ d\end{bmatrix} =\begin{bmatrix}c^2\Delta T_{12}^2+x_1^2+y_1^2-x_2^2-y_2^2\\ c^2\Delta T_{13}^2+x_1^2+y_1^2-x_3^2-y_3^2\\ c^2\Delta T_{14}^2+x_1^2+y_1^2-x_4^2-y_4^2\end{bmatrix}. \tag{3}$$

The source location (u,v) and the distance between sensor 1 and the source (d) can then be solved for by multiplying both sides of Eq. (3) by the inverse of the matrix, resulting in Eq. (4):

$$\begin{bmatrix}u\\ v\\ d\end{bmatrix}=\begin{bmatrix}2x_1-2x_2 & 2y_1-2y_2 & -2c\Delta T_{12}^2\\ 2x_1-2x_3 & 2y_1-2y_3 & -2c\Delta T_{13}^2\\ 2x_1-2x_4 & 2y_1-2y_4 & -2c\Delta T_{14}^2\end{bmatrix}^{-1} \times\begin{bmatrix}c^2\Delta T_{12}^2+x_1^2+y_1^2-x_2^2-y_2^2\\ c^2\Delta T_{13}^2+x_1^2+y_1^2-x_3^2-y_3^2\\ c^2\Delta T_{14}^2+x_1^2+y_1^2-x_4^2-y_4^2\end{bmatrix}. \tag{4}$$

The time-of-flight differences (ΔT_{1j}) were determined by selecting characteristic points in the manatee vocalization and then comparing the times at which that point occurred at each of the hydrophones. The time-of-flight differences can then be substituted into Eq. (4) and u, v, and d can be determined.

B. Hyperbolic fixing

The second approach to locating the position of the source was performed using the program Ishmael (Mellinger, 2001). Hyperbolic fixing is the method chosen for position location in this approach. Hyperbolic fixing also uses time-of-flight differences to determine the location; however, it can rely on as little as three sensors to estimate the position of an unknown source in a plane. The following is an example of two-dimensional hyperbolic fixing, based on the three-dimensional case presented by Bertrand Fang (Fang, 1990). The time-of-flight differences (t_{1j}) between sensor 1 and the remaining two are defined in Eq. (5):

$$\begin{aligned}t_{12}&=t_1-t_2,\\ t_{13}&=t_1-t_3.\end{aligned} \tag{5}$$

To simplify the equations the sensors are placed at specific locations in the coordinate system: sensor 1 at (0,0), sensor 2 at $(x_2,0)$, and sensor 3 at (x_3,y_3). The unknown position is located at (u,v). The distance between the sensors and the unknown location is given by Eq. (6), where c is the speed of sound and R_{1j} is the difference in distance from the source to sensor 1 and the remaining two sensors:

$$\begin{aligned}\sqrt{u^2+v^2}-\sqrt{(u-x_2)^2+v^2}&=c\cdot t_{12}=R_{12},\\ \sqrt{u^2+v^2}-\sqrt{(u-x_3)^2+(v-y_3)^2}&=c\cdot t_{13}=R_{13}.\end{aligned} \tag{6}$$

Equation (7) is derived by squaring both sides of Eq. (6) and rearranging the terms:

$$\begin{aligned}\frac{R_{12}^2-x_2^2+2x_2x}{2R_{12}}&=\sqrt{u^2+v^2},\\ \frac{R_{13}^2-x_3^2-y_3^2+2x_3x+2y_3y}{2R_{13}}&=\sqrt{u^2+v^2}.\end{aligned} \tag{7}$$

The first equation in Eq. (7) is then substituted into the second and then solved for v in terms of u, resulting in Eq. (8),

$$v=g\cdot u+h, \tag{8}$$

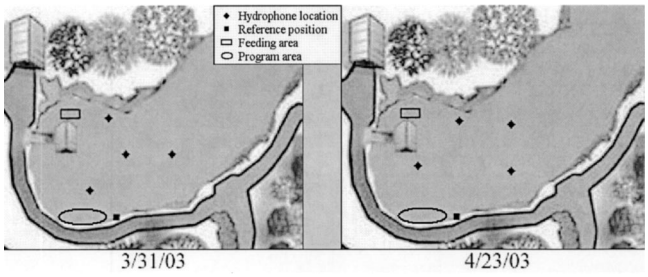


FIG. 1. Hydrophone locations in Homosassa Springs (top view).

where

$$g=\frac{R_{13}x_2/R_{12}-x_3}{y_3},$$

$$h=\frac{x_3^2+y_3^2-R_{13}^2+R_{13}R_{12}(1-(x_2/R_{12})^2)}{2y_3}.$$

The unknown value u can then be obtained by substituting Eq. (8) into Eq. (7) resulting in Eq. (9),

$$d\cdot u^2+e\cdot u+f=0, \tag{9}$$

where

$$d=\left(\frac{x_2}{R_{12}}\right)^2-g^2-1,$$

$$e=x_2\left(1-\left(\frac{x_2}{R_{12}}\right)^2\right)-2g\cdot h,$$

$$f=\frac{R_{12}^2}{4}\left(1-\left(\frac{x_2}{R_{12}}\right)^2\right)^2-h^2.$$

Equation (9) can then be solved for u and then used in Eq. (8) to determine v. Equation (9) will lead to two possible solutions. A single solution can be obtained by either ruling out one of the solutions as an impossibility or by using a forth sensor to eliminate the ambiguity. The number of solutions was reduced to one, because four hydrophones were used during the recording process.

Ishmael uses a cross-correlation method to determine the time-of-flight differences (t_{1j}). The hydrophone signals from sensors 2–4 are each cross-correlated with hydrophone signal 1. The time-delay which results in the maximum cross-correlation is taken to be the time-of-flight difference.

III. EXPERIMENTAL DESCRIPTION

Approximately 11 h of manatee vocalizations were recorded in Homosassa Springs Wildlife State Park on 31 March and 23 April 2003. Homosassa Springs contains nine female manatees and no male manatees. The manatees in Homosassa Springs range in age from 9 to 37 years. They can not leave the park and wild manatees can not enter the park due to a physical barrier in the spring run that prohibits their interaction.

For each test, four hydrophones were placed within the spring, as shown in Fig. 1, and recorded using a Teac RD-135T DAT recorder with a sample rate of 48 kHz. A pole at the edge of the spring was used as a reference point and the distances to each hydrophone were recorded. The distance between the hydrophones was also recorded, so that the relative positions could then be calculated.

At different times during the day a Lubell LL-98 underwater speaker was used to broadcast different calls to the manatees. Since a controlled sound exposure for manatees has never been reported in the literature, the authors believed that the broadcast test was worth performing. A diver would place the speaker in the spring, and before broadcasting any sounds, the diver would swim away and the manatees were allowed time to see and inspect the speaker. This allowed the researchers to observe the manatees' response to the sounds which were broadcast, while minimizing the manatees' response from visual stimulation. For a period of at least 15 min prior to broadcast, the manatees showed no interest in the speaker. The speaker was moved between broadcast sessions so that the manatees would not become accustomed to the broadcasts being produced at one specific location within the spring.

Several events occur in Homosassa Springs Wildlife State Park daily. Several manatee programs take place in which a person enters the water and discusses the manatees. During the program, the nine manatees are in the vicinity of the oval shown in Fig. 1 and are fed carrots and vitamins. There is also a feeding area, represented by the black rectangle, and a manatee show. Twice daily, the manatees are fed a mixture of vegetables here. The broadcasting times were chosen so that they would not overlap with either the manatee programs or the feeding times. There is also a floating observation center in the center of the spring located next to the feeding area.

The recordings were later replayed in the laboratory and the number of audible manatee vocalizations was determined for every 5-min period. Knowing that there are nine manatees present in the spring at all times, and knowing the total number of vocalizations, the average vocalization rate per manatee can be determined.

IV. CALIBRATION

Four High Tech, Inc. (HTI) hydrophones were used during the recording of the manatee vocalizations: two model HTI-90-U hydrophones and two model HTI-96-MIN hydrophones. The four HTI hydrophones were recently purchased and calibrated by the manufacturer. The manufacturer's calibration was verified in a pool and in the Homosassa River.

First, the manufacturer's calibration was verified by comparing their voltage output with that of two B&K 8104 hydrophones. A 1-kHz tone was generated with a Lubell LL-98 underwater loudspeaker in a 50-m Olympic pool. Each individual hydrophone was mounted in a stand sequentially, approximately 10 m from the loudspeaker. The outputs from the two B&K hydrophones were then averaged and used to determine the sensitivity of the four HTI hydrophones. Each hydrophone was tested separately and placed in the same location as the B&K hydrophone (± 0.005 m), for a fixed source level from the speaker. No significant variations in the sound levels were detected within ± 0.5 m of the measurement location, as the hydrophones were moved about the calibration location.

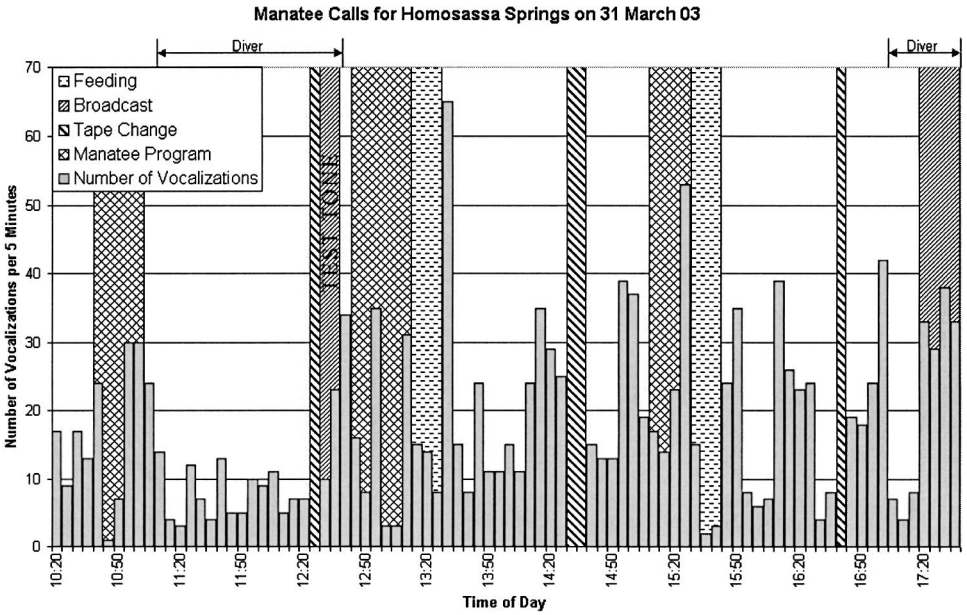


FIG. 2. Number of manatees vocalizations at Homosassa Springs on 31 March 2003.

Second, the calibration was verified using a swept-sine broadcast in a spring at Homosassa River. The swept-sine broadcast was performed between 0.3 and 10 kHz with the hydrophones placed approximately 25 m away from the source. For this test, the hydrophones were tested simultaneously and were collocated. The separation between hydrophones was approximately 20 cm. B&K 8104 hydrophones were used as the comparison hydrophones. The B&K 8104 hydrophones have a flat response over this range and have been calibrated by the manufacturer. The calibration of the B&K hydrophones was verified using a B&K Type 4229 pistonphone, with an accuracy of $\pm 0.1\%$. The two calibration tests demonstrated that the four HTI hydrophones were within ± 1 dB of the manufacturer's specifications.

V. RESULTS

A. Homosassa Springs—31 March 2003

At Homosassa Springs on 31 March 2003, approximately 7 h of manatee vocalizations were recorded. The recordings started at 10:20 a.m. and continued until 5:40 p.m. The number of manatee vocalizations heard on the recording was then counted for every 5-min period (see Fig. 2). In addition to the manatee program and the feeding times, the times during tape changes and broadcast periods are marked. Several times during the day a diver was also present in the spring. These times are also labeled along the top of Fig. 2.

On this date, two different sets of broadcasts were performed. Between 12:30 p.m. and 12:40 p.m. a 5-kHz harmonic test tone was broadcast at low levels to verify that the speaker and all four hydrophones were operational. The second broadcast involved manatee vocalizations that had previously been recorded at Blue Spring State Park. Approximately 50 to 100 manatees were in the vicinity of Blue Spring the day the authors' recordings were made. Because of the large number of animals present at the time of recording, the vocalizations could be any variety of manatee calls. These recorded calls were broadcast between 5:20 p.m. and 5:40 p.m.

During the day when no broadcasting was taking place, it was observed that the manatees vocalized at an average rate of 1.88 calls/manatee/5 min (see Table I). There was negligible change (1.83 calls/manatee/5 min) in the manatees' vocalization rate in response to the speaker test tone signal. When the Blue Spring recordings were broadcast in Homosassa Springs, the manatee vocalization rate doubled (3.69 calls/manatee/5 min). However, it should be noted that there were some 5-min periods in which the vocalization rates were just as high. This typically occurred when Homosassa park program was transpiring and the manatees were receiving carrots during the show. It is important to point out that the influence of the diver being present with the manatees was not investigated. These particular manatees are very accustomed to divers since divers are present in the spring

TABLE I. Manatee vocalization rates on 31 March 2003 at Homosassa Springs.

	Excluding broadcasting	5-kHz test tone	Broadcast of Blue Springs calls
Average no. of calls/5 min	16.88	16.50	33.25
Total time recorded (min)	390	10	20
Total no. of calls	1317	33	133
Average no. of calls/manatee/5 min	1.88	1.83	3.69
Standard deviation	1.40	1.02	0.41

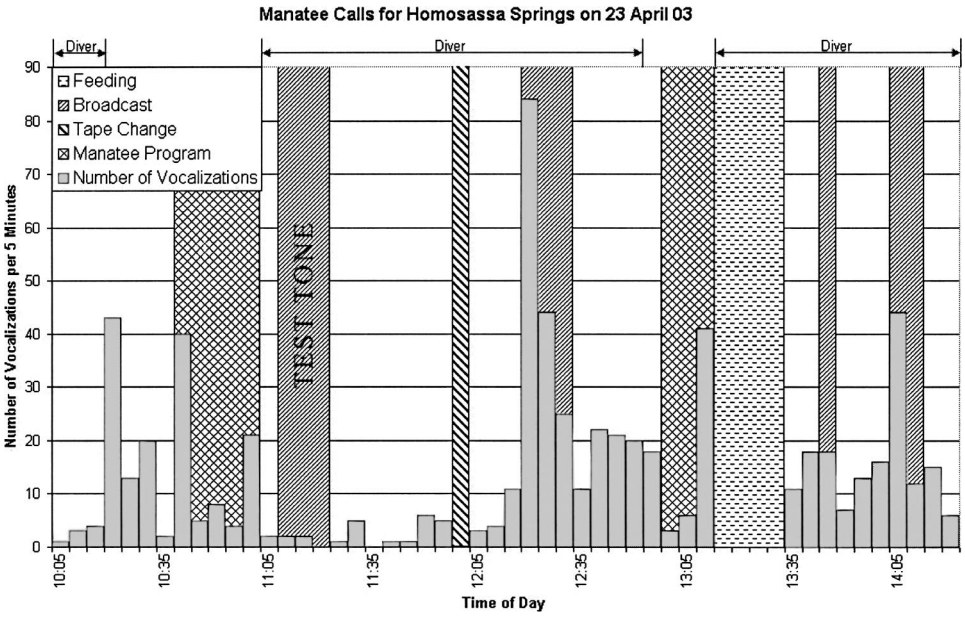


FIG. 3. Number of manatees vocalizations at Homosassa Springs on 23 April 2003.

run on a daily basis to clean the windows of the viewing area and to conduct research.

Prior to the broadcast (vocalizations recorded at Blue Spring), the manatees ignored the speaker and did not appear to be interested in its presence. Once the broadcasting began, six of the nine manatees converged onto the speaker to investigate the sound. The manatees were extremely active during the broadcast and congregated to within a few meters of the speaker. It was obvious to the researchers that the manatees were well aware of the broadcast. Once the broadcast ended the manatees dissipated and resumed their normal activity. The three manatees that did not respond were located at the other end of the spring run and it was likely that they could not hear the broadcast signal. This response was not observed during the speaker test or at any other time during the day.

B. Homosassa Springs—23 April 2003

Another four hours of manatee vocalizations were recorded at Homosassa Springs on 23 April 2003 between 10:05 a.m. and 2:25 p.m. and four different broadcasts were performed (see Fig. 3). The speaker test was performed at 11:10 a.m. and the Blue Spring calls were broadcast at 12:20 p.m. The final two broadcasts used recordings made by O'Shea (O'Shea, 1981–1984). The third broadcast period (1:45 p.m.) used recordings that O'Shea describes as a frightened calf and the final broadcast period (2:05 p.m.) used various manatee calls recorded by O'Shea at 1 m from the manatee.

The average vocalization rate excluding broadcast periods was 1.09 calls/manatee/5 min (see Table II). The speaker test tone was performed shortly after the manatees finished feeding and was at the quietest part of the day. The manatees again ignored the sounds of the speaker test tone and averaged only 0.15 calls/manatee/5 min. The average manatee vocalization rate during the broadcast (recorded at Blue Spring) session increased to approximately five times the rate when no broadcasts were occurring, 5.67 calls/manatee/5 min. When the calls of a frightened calf were broadcast to the female manatees, the vocalization rate was 2.00 calls/manatee/5 min and the various 1-m recordings produced vocalization rate of 3.11 calls/manatee/5 min.

For each of the broadcasts, the manatees reacted in a similar manner as described for the test on 31 March. When the speaker test tone was broadcast the manatees showed no interest in the speaker. However, when the three sets of prerecorded manatee vocalizations were broadcast, five, six, and eight manatees congregated to within a few meters of the speaker for the second, third, and forth broadcasts, respectively. When the distressed calf vocalizations were broadcast the female manatees encircled the speaker with their tails inside the circle and their heads looking outward.

TABLE II. Manatee vocalization rates on 23 April 2003 at Homosassa Springs.

	Excluding broadcasting	5-kHz test tone	Broadcast of Blue Springs calls	Broadcast of frightened calf calls	Broadcast of misc. O'Shea calls
Avgerage no. of calls/5 min	10.26	1.33	51.00	18.00	28.00
Total time recorded (min)	210	10	15	5	10
Total no. of calls	431	4	153	18	56
Avgerage no. of calls/manatee/5 min	1.14	0.15	5.67	2.00	3.11
Standard deviation	1.24	0.13	3.35	N/A	2.51

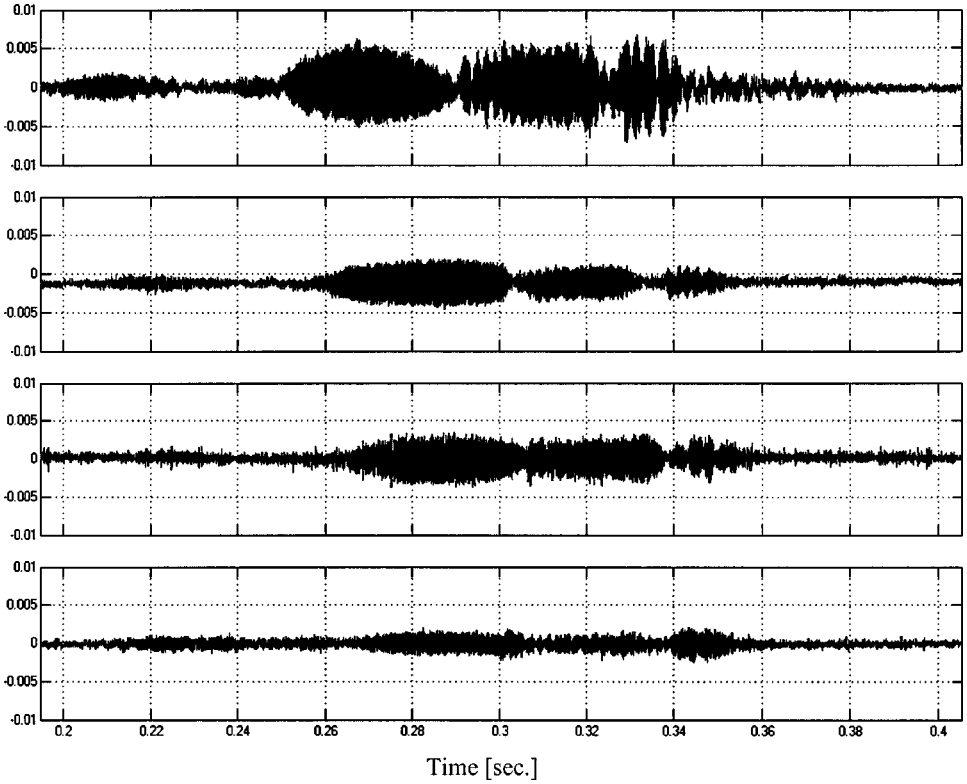


FIG. 4. Sample manatee call on four hydrophones.

C. Triangulation and source level

The manatee vocalizations are triangulated using the two methods previously described, and the estimated positions from each of the methods are compared. The estimated location is considered to be accurate when the position from Ishmael and from Mahajan's method differed by less than 3 m. Because Ishmael uses the cross-correlation between the signals and not a point comparison, the estimated position given by Ishmael was used as the location of the manatee.

The original recordings (see Fig. 4) were passed through a 500-Hz high-pass filter to reduce any low-frequency background noise. The fundamental frequency of manatee vocalizations is typically between 2 and 5 kHz, so the high-pass filter will not significantly alter the spectral content of the manatee vocalizations. The signals were then placed into buffers with a length of 1024 and the rms pressure was then calculated. The maximum sound pressure level during the manatee vocalization was recorded as the peak level.

The source level is the calculated sound pressure level at 1 m from the manatee's estimated position based upon each of the four hydrophones. The results from 50 manatee vocalizations are shown in Fig. 5. Because each of the 50 manatee vocalizations was recorded on four hydrophones and each hydrophone is being used to determine the source level, 200 results are presented. Since the majority of the spring is less than 2 m in depth, the source level was based upon the assumption of cylindrical spreading. The mean source level for the 50 manatee vocalizations was found to be 112.5 dB. All sound pressure levels are based upon the standard underwater reference pressure of 1 μPa.

The mean source level for each vocalization was determined by averaging the source rms pressure values calculated from each of the four hydrophones. The average source level values for the 50 manatee vocalizations are presented in Fig. 6. Based upon the average values for each vocalization, the mean source level for the 50 manatee vocalizations was found to be 113.2 dB.

The authors performed this research with the assumption that the manatee vocalizations can be represented as sources with no directivity in an environment that can be characterized by a cylindrical spreading model. While much of the data fit this model within a few decibels, there were instances where the source level estimates deviated significantly (up to 22.8 dB). The mean difference between the highest and lowest source level calculated using two different hydrophone measurements is 8.6 dB. To verify that this difference is not attributed to the sensitivity of any particular hydrophone being offset, the values for the source level estimates were averaged for each hydrophone. The average source level

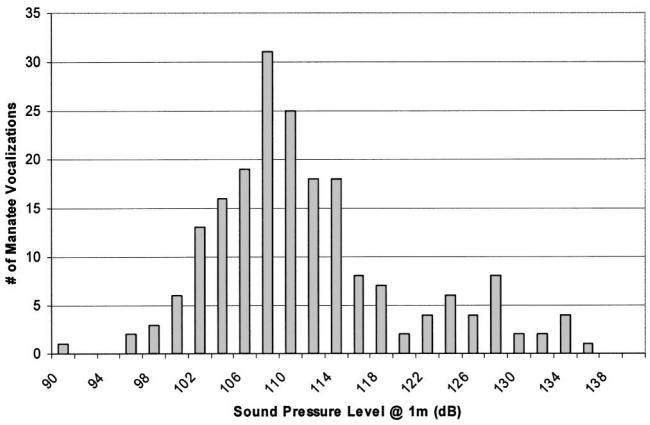


FIG. 5. Source level for 50 manatee vocalizations received on four hydrophones.

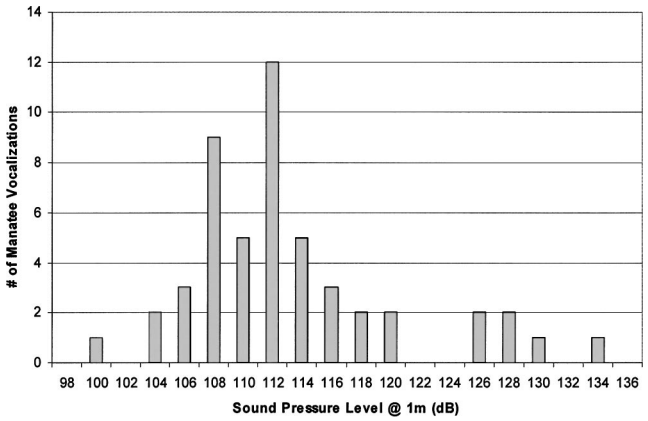


FIG. 6. Average source level for 50 manatee vocalizations.

value for each hydrophone was within ±1.25 dB of each other. These observations led the authors to believe that other factors may be involved in the acoustic spreading of the manatee vocalizations. The spring geometry may have acoustic mode shapes that will affect the acoustic levels within the spring. Another possibility in the measured differences may be attributed by the directivity of the manatee vocalization itself.

VI. CONCLUSIONS

Within this paper, the vocalization rates of the nine female West Indian manatees present in Homosassa Springs Wildlife State Park were recorded on 31 March and 23 April 2003. The average vocalization rates were determined to be 1.83 and 1.09 calls/manatee/5 min. These results are consistent with previous studies performed with the West Indian manatee.

The manatees' response to several broadcast calls was observed. On both occasions when the recorded Blue Spring manatee vocalizations were broadcast into Homosassa Springs, the vocalizations rates increased (approximately two to five times the average nonbroadcast rates). The distressed calf recording and the various O'Shea recordings evoked similar vocalization responses in the Homosassa manatees (approximately two to three times the average nonbroadcast rates). In addition to responding vocally to the broadcasts the manatees responded physically. During all broadcast sessions between five and eight of the nine manatees present congregated within a few meters of the speaker. The physical responses of the manatees varied between the different broadcasts. When the distressed calf vocalizations were broadcast the female manatees encircled the speaker with their tails inside the circle and their heads looking outward. When the recorded Blue Spring vocalizations were broadcast, the manatees approached the speaker and became increasingly active. The manatees started splashing in the water and repeatedly rolled above the speaker.

The manatees' responses to the broadcasted vocalizations show that they can detect the source location of the call. Also, the responses to the broadcast calls led the authors to believe that the vocalizations are clearly a form of communication for the manatee. The manatees displayed an increased vocalization rate when unfamiliar calls were broadcast. Different physical responses were observed for the various manatee vocalizations.

The mean source level of the manatees' vocalizations was determined based upon 50 recorded and triangulated manatee vocalizations. The mean source level was determined using two different methods. The mean source level using each hydrophone separately was determined to be 112.5 dB. When the rms pressure from the four hydrophones was averaged for each vocalization the mean source level was determined to be 113.2 dB.

Ideally the source level determined from the four hydrophones would be nearly identical, if the sound source had no directivity. Because this was not found to be true, the manatees probably have some directivity in their vocalizations, however this remains to be verified. It is also likely that the discrepancies are due to the geometry and geology of the spring. Homosassa Spring is barren of vegetation and has a sandy bottom; however numerous large rocks are present throughout the spring area that could generate acoustic modes and cause a deviation from the cylindrical spreading model used.

ACKNOWLEDGMENTS

The authors would like to express their sincere appreciation to the Florida Fish and Wildlife Conservation Commission in supporting this research.

Bengston, J. L., and Fitzgerald, S. M. (**1985**). ''Potential Role of vocalizations in West Indian manatees,'' J. Mammal. **66**(4), 816–819.

Fang, B. T. (**1990**). ''Simple Solutions for Hyperbolic and Related Position Fixes,'' IEEE Trans. Aerosp. Electron. Syst. **26**(5), 748–753.

Florida Department of Environmental Protection, Division of Marine Resources (**1996**). Save the Manatee Trust Fund, Fiscal Year 1995–1996, Annual Report, Florida Marine Research Institute, 100 Eighth Avenue S.E., St. Petersburg, FL 33701-5095.

Florida Fish and Wildlife Conservation Commission (**2002**). Save the Manatee Trust Fund, Fiscal Year 2001–2002, Annual Report, Florida Fish and Wildlife Conservation Commission, 620 South Meridian Street, OES-BPS, Tallahassee, FL 32399-1600.

Mahajan, A., and Walworth, M. (**2001**). ''3-D Position Sensing Using the Differences in the Time-of-Flights from a Wave Source to Various Receivers,'' IEEE Trans. Rob. Autom. **17**(1), 91–94.

Mellinger, D. K. (**2001**). Ishmael 1.0 User's Guide. NOAA, Technical Memorandum OAR PMEL-120, NOAA/PMEL, 7600, Sand Point Way NE, Seattle, WA 98115-6349.

Niezrecki, C., Phillips, R., Meyer, M., and Beusse, D. O. (**2003**). ''Acoustic detection of manatee vocalizations,'' J. Acoust. Soc. Am. **114**, 1640–1647.

Nowacek, D. P., Casper, B. M., Wells, R. S., Nowacek, S. M., and Mann, D. A. (**2003**). ''Intraspecific and geographic variation of West Indian Manatee,'' J. Acoust. Soc. Am. **114**, 66–69.

O'Shea, T. (**1981–1984**). ''Manatee vocalization-catalog of sounds,'' produced by Coastal Systems Station, Naval Surface Warfare Center, Dahlgren Division, Panama City, FL.